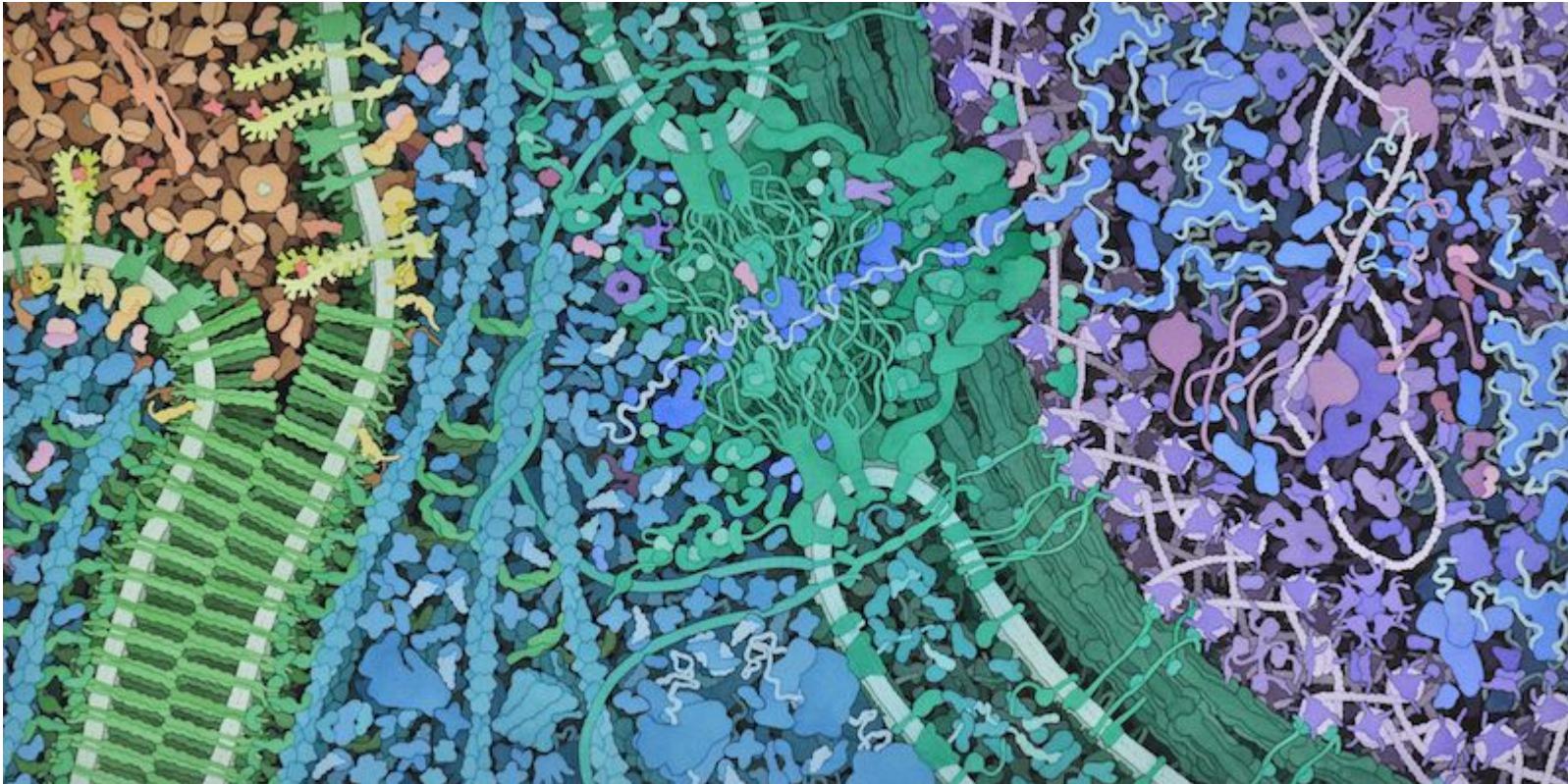


AMIDD Lecture 7: From individual interactions to networks



Vascular Endothelial Growth Factor (VegF) Signaling, David S. Goodsell, 2011

Dr. Jitao David Zhang, Computational Biologist

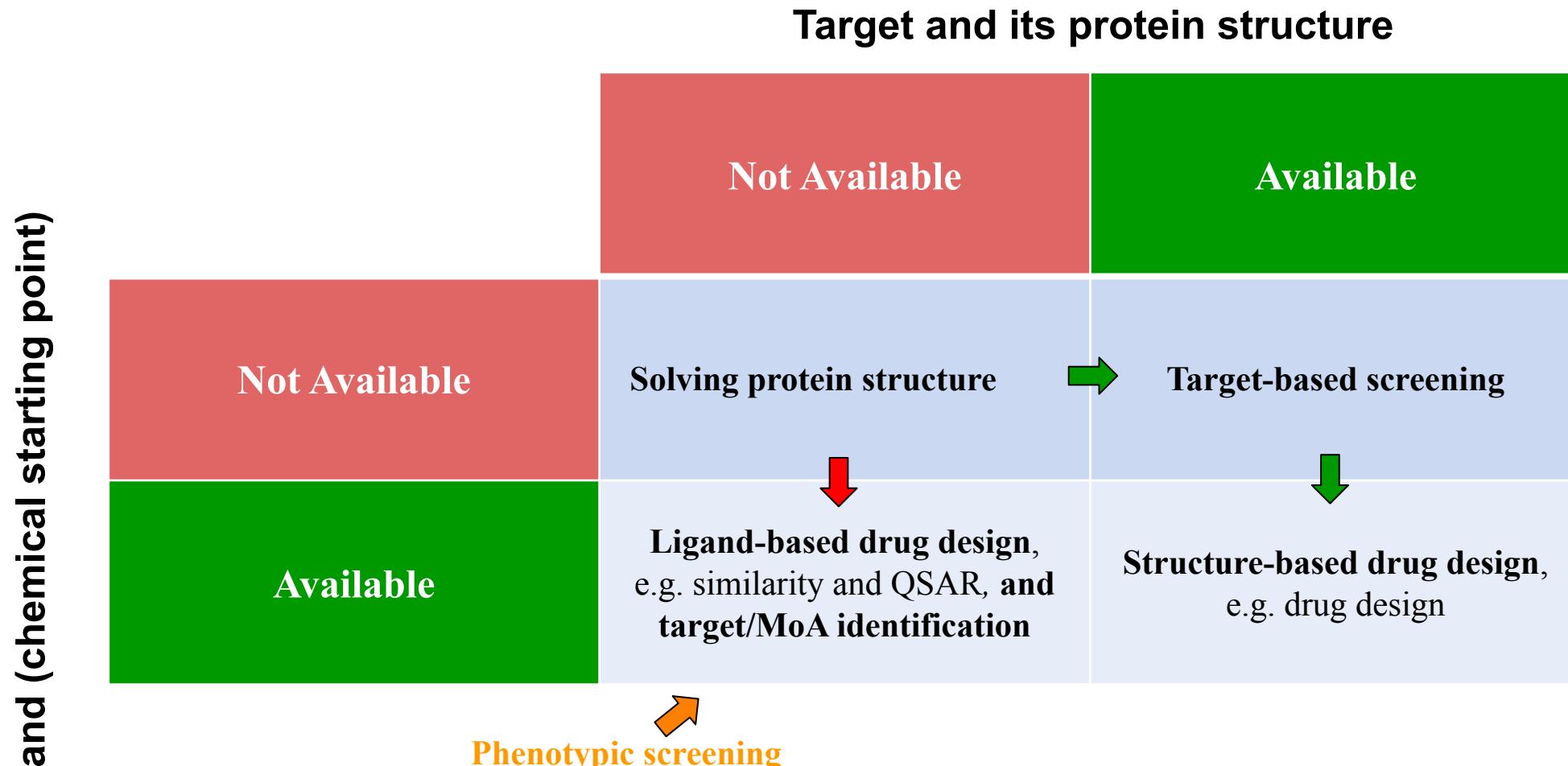
¹ Pharmaceutical Sciences, Pharma Research and Early Development, Roche Innovation Center Basel, F. Hoffmann-La Roche

² Department of Mathematics and Informatics, University of Basel

Topics

- QSAR, machine learning, and causal inference
- Drug-target interaction: biophysical and biochemical views
- Biological networks interact with drugs and manifest its efficacy and safety

Structure-based and ligand-based drug design



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

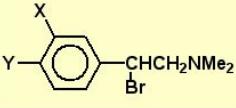
Quantitative Structure-Activity Relationships (QSARs)

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

Compounds (C)	Target property	Molecular Descriptors (MD)					
		MD ₁	MD ₂	...	MD _M		
C ₁	y ₁	x _{1,1}	x _{1,2}	...	x _{1,M}		
C ₂	y ₂	x _{2,1}		
C ₃	y ₃		
C ₄	y ₄		
...		
...		
C _N	y _N	x _{N,1}	x _{N,2}	...	x _{N,M}		

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.

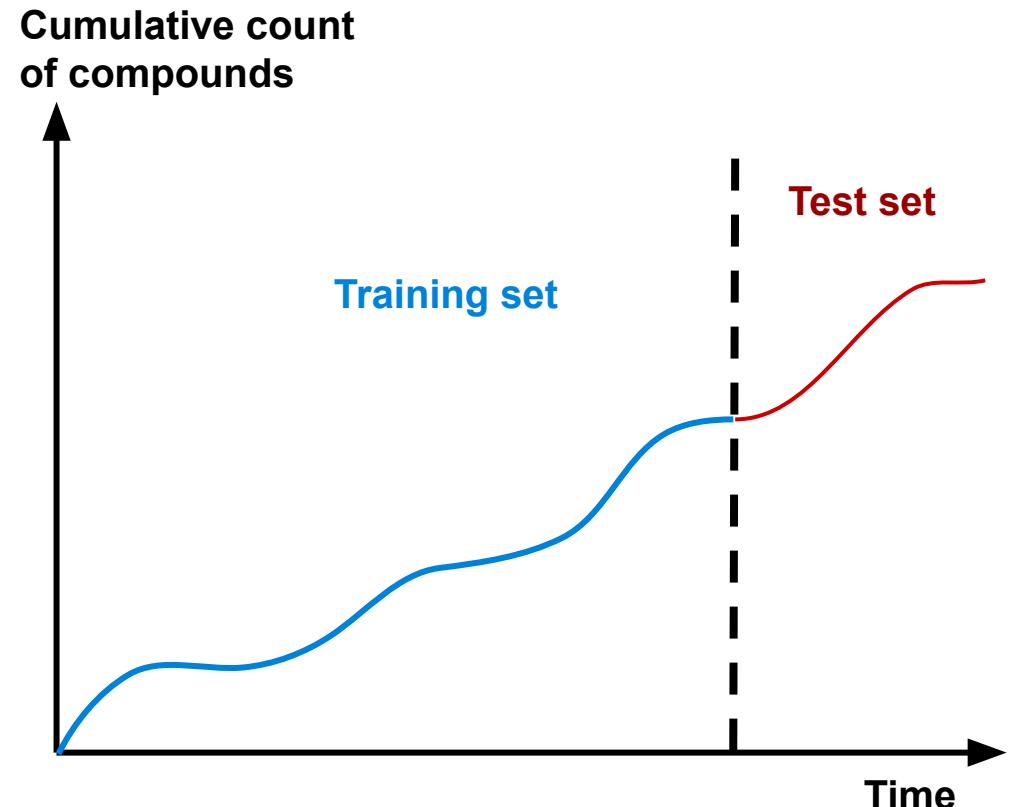
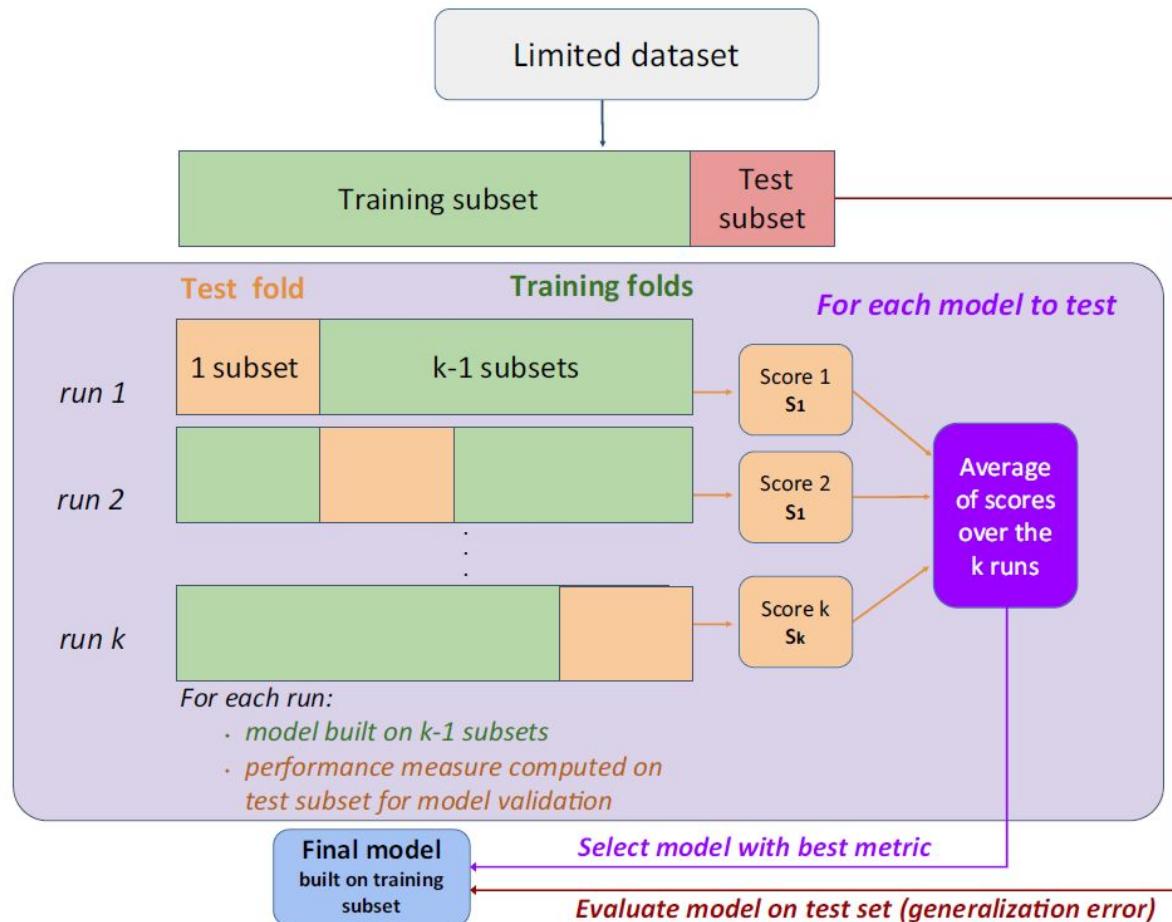


meta	para	meta-	para-						log 1/C	log 1/C				
(X)	(Y)		F	Cl	Br	I	Me	F	Cl	Br	I	Me	obsd.	calc.a)
H	H												7.46	7.82
H	F							1					8.16	8.16
H	Cl								1				8.68	8.59
H	Br									1			8.89	8.84
H	I										1		9.25	9.25
H	Me											1	9.30	9.08
F	H		1										7.52	7.52
Cl	H			1									8.16	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					
Cl	Cl				1						1			
Br	Cl					1						1		
Me	Cl						1					1		
Cl	Br				1									
Br	Br					1								
Me	Br							1						
Me	Me								1					
Br	Me									1				

Multivariate regression analysis

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

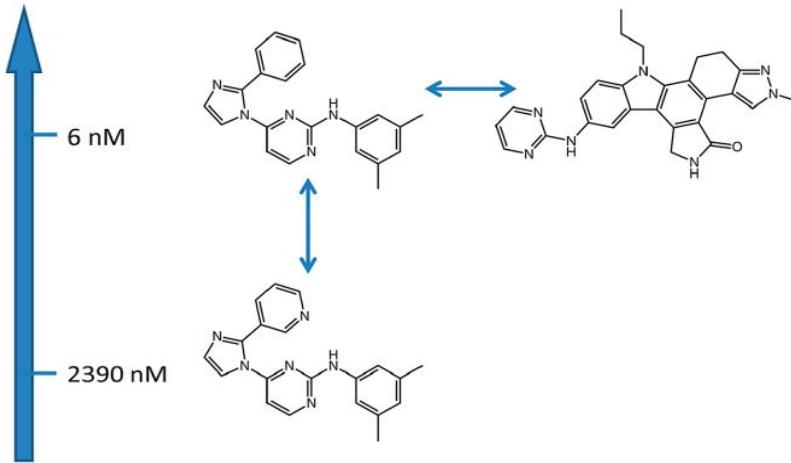
Watchout 1: Temporal validation is essential for drug discovery



(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.

(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.

Watchout 2: Molecular similarity does not equal biological similarity



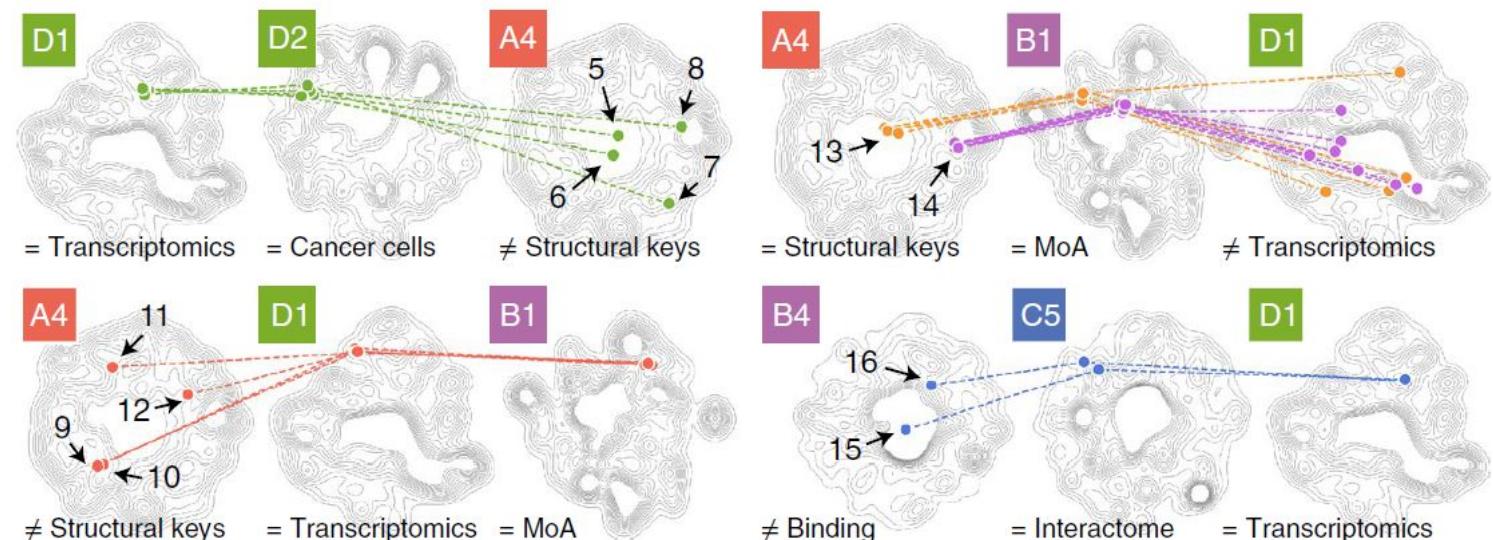
Watch out biological activity cliffs!

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
A	Red	Red	Red	Red	Red
B	Purple	Purple	Purple	Purple	Purple
C	Blue	Blue	Blue	Blue	Blue
D	Green	Green	Green	Green	Green
E	Orange	Orange	Orange	Orange	Orange

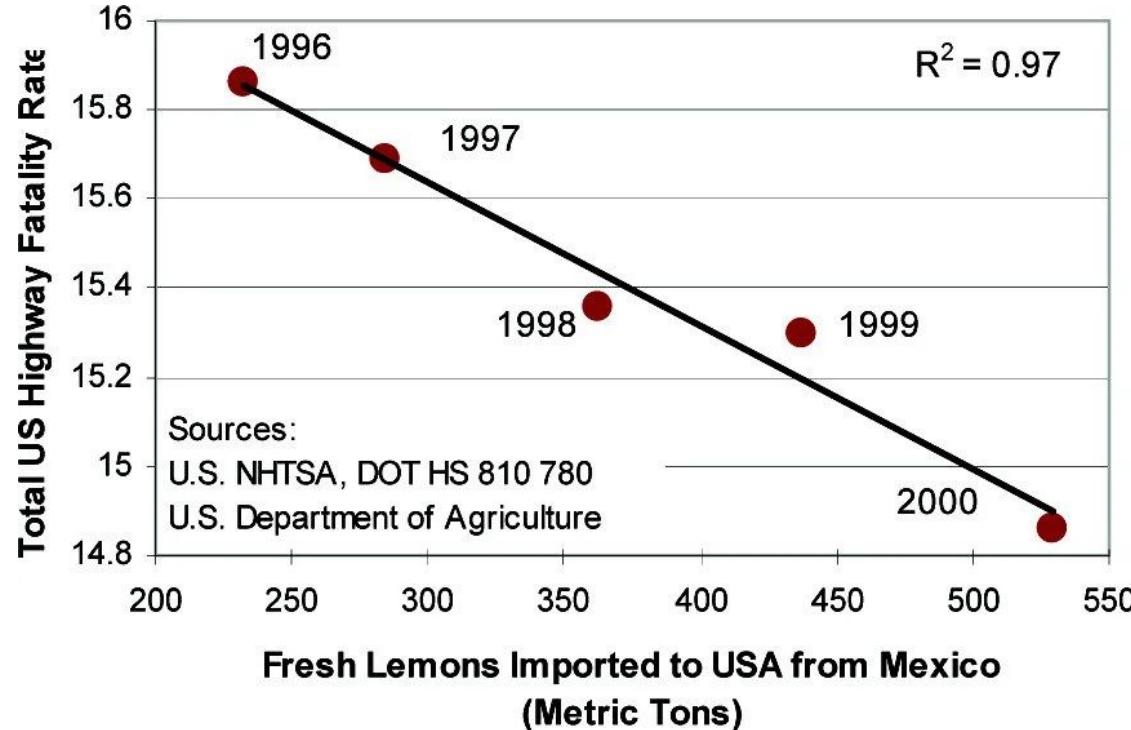
A1: 2D fingerprints
 B1: Mechanisms of action
 C1: Small molecule roles
 D1: Transcription
 E1: Therapeutic areas
 A2: 3D fingerprints
 B2: Metabolic genes
 C2: Small molecule pathways
 D2: Cancer cell lines
 E2: Indications
 A3: Screens
 B3: Crystals
 C3: Signaling pathways
 D3: Chemical genetics
 E3: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

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



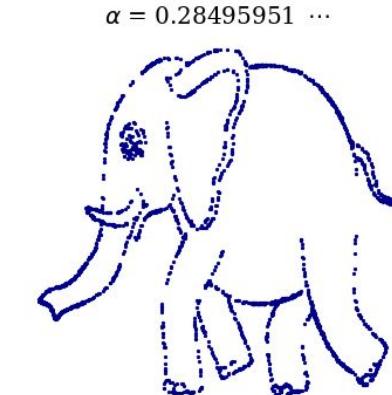
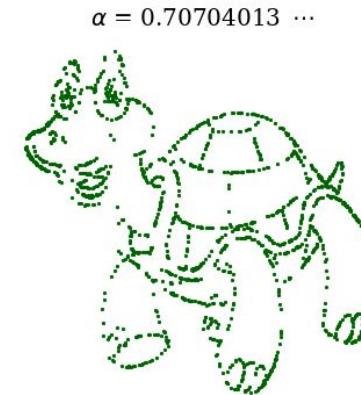
Duran-Frigola, Miquel, Eduardo Pauls, Oriol Guitart-Pla, Martino Bertoni, Víctor Alcalde, David Amat, Teresa Juan-Blanco, and Patrick Aloy. 2020. “[Extending the Small-Molecule Similarity Principle to All Levels of Biology with the Chemical Checker](#).” Nature Biotechnology, May, 1–10.

Watchout 3: Correlation can be caused by causation, confounding, coincidence, and conspiracy. Do we need correlation or causation?



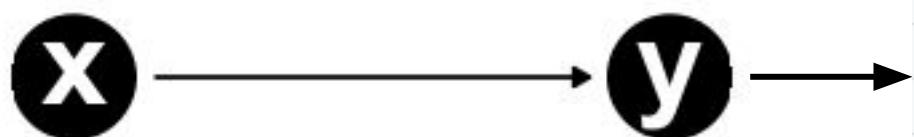
Johnson, Stephen R. "The Trouble with QSAR (or How I Learned To Stop Worrying and Embrace Fallacy)." *Journal of Chemical Information and Modeling* 48, no. 1 (January 1, 2008): 25–26.
<https://doi.org/10.1021/ci700332k>

$$f_{\alpha}(x) = \sin^2(2^{x\tau} \arcsin \sqrt{\alpha})$$



Boué, Laurent. "Real Numbers, Data Science and Chaos: How to Fit Any Dataset with a Single Parameter." ArXiv:1904.12320 [Cs, Stat], April 28, 2019. <http://arxiv.org/abs/1904.12320.GitHub Repo>.
 Also see: [Drawing an elephant with four complex parameters](#)

Generative models shed light on correlation and causality

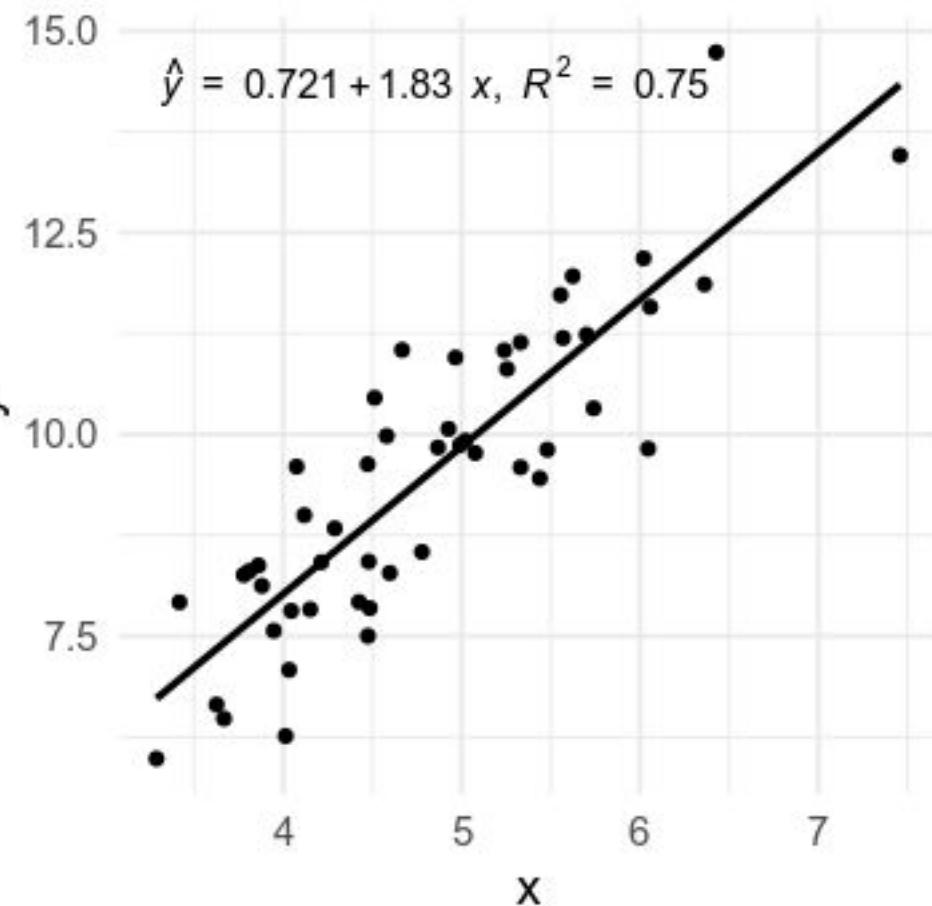


	x	y
1	4.926791	10.067779
2	4.479734	8.424283
3	4.289686	8.835629
4	4.474023	9.630499
5	4.214551	8.416680
6	6.057431	11.578080
7	4.597903	8.283025
8	5.021571	9.922731
9	3.627323	6.651222
10	5.622794	11.959972
11	5.555025	11.727815
12	4.966007	10.951562
13	5.076791	9.768299

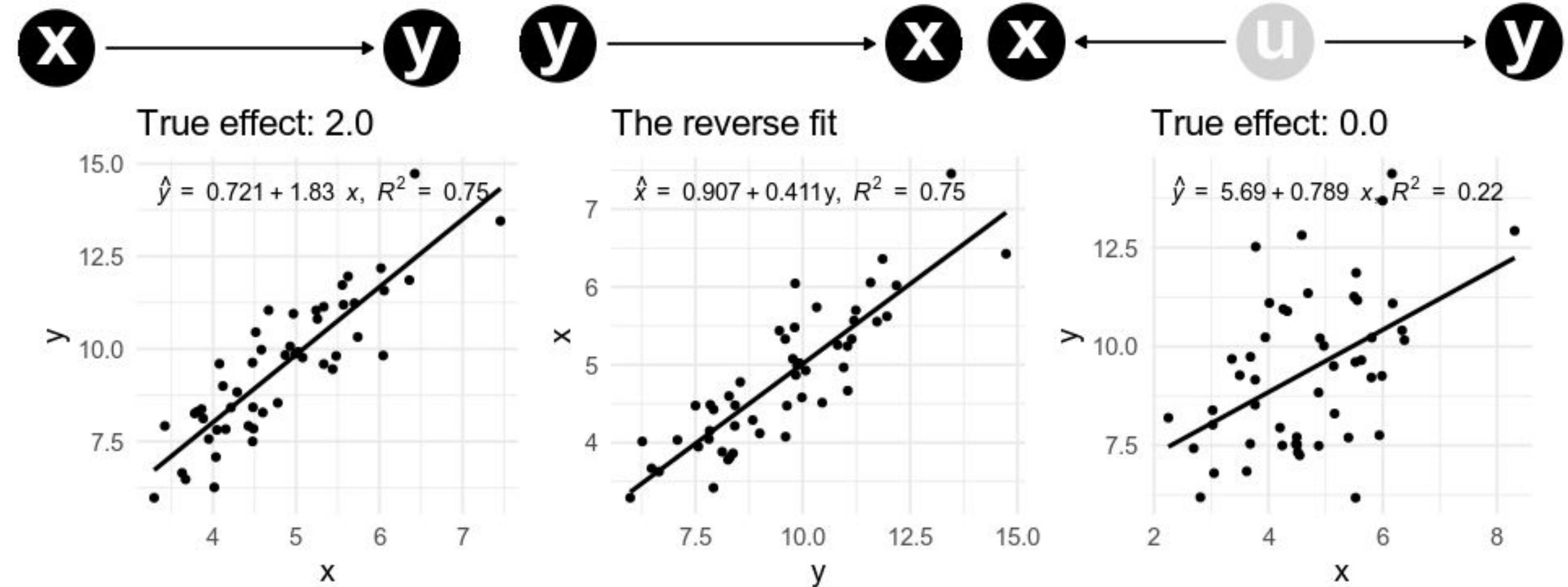
Assumptions of the **generative model**:

1. X is a random variable;
2. Every unit change of X increases a change of 2 units in Y.

True effect: 2.0



Correlation is caused by causation or confounding



Statistical models alone cannot derive causality from correlation

We learn causality by (1) listing models explicitly and (2) manipulating a variable and observe the outcomes

Model 1



Model 2

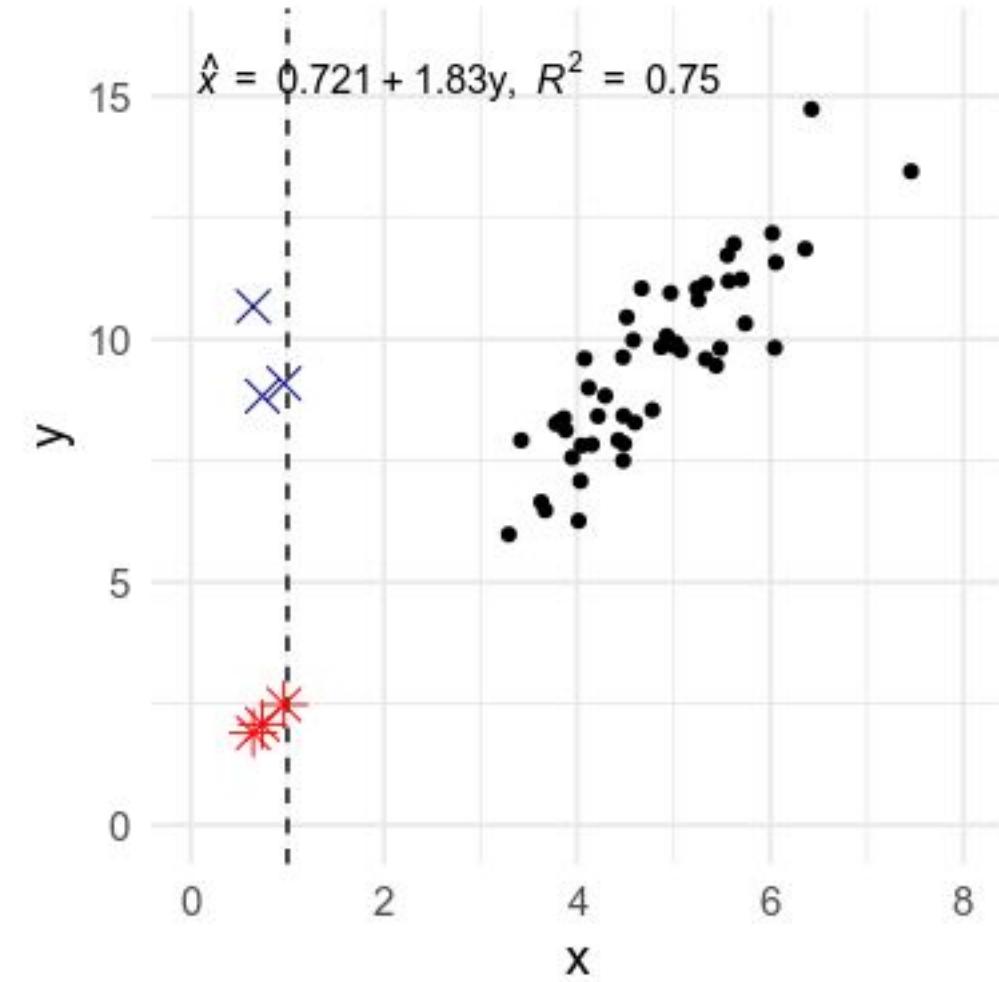


Model 3



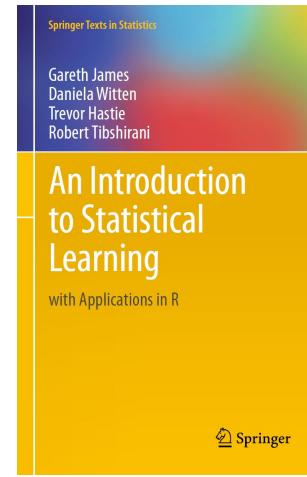
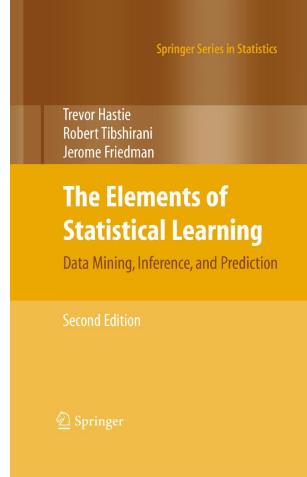
Assume that the data is generated by either Model 1, or Model 2, or Model 3. And assume that we can manipulate the value of X by setting it to 1.0 (the dash line).

**Question: which outcomes (red stars or blue crosses) would support which models?
Why?**

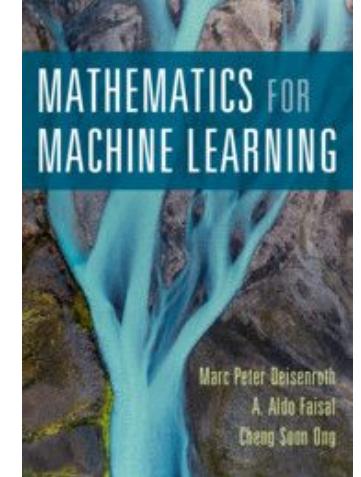


Resources for learning about machine learning

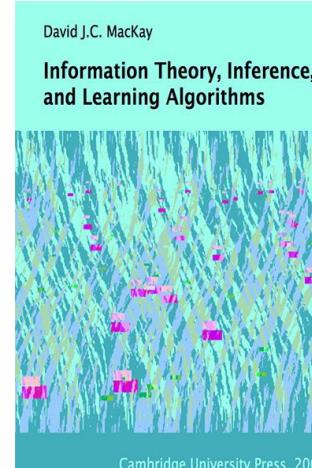
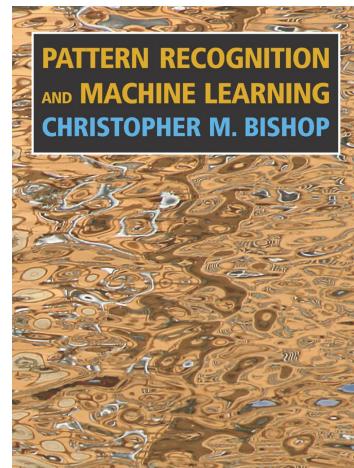
[ESL](#) and [ISL](#): From a frequentist view (almost)



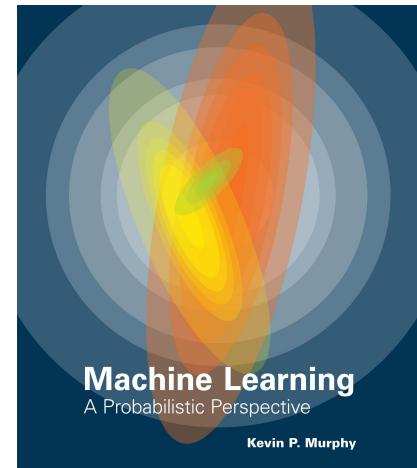
[Mathematical foundations](#)



[PRML](#) and [ITILA](#): From a Bayesian view

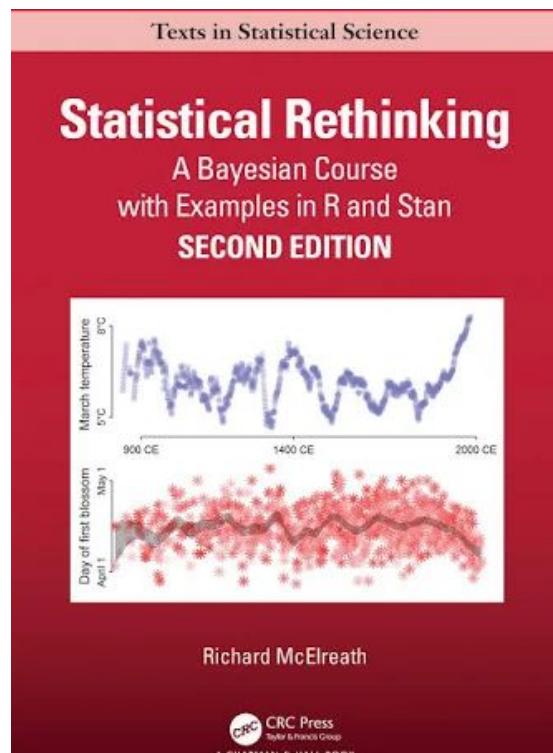


[MLaPP](#): Application oriented, more accessible, and balanced views

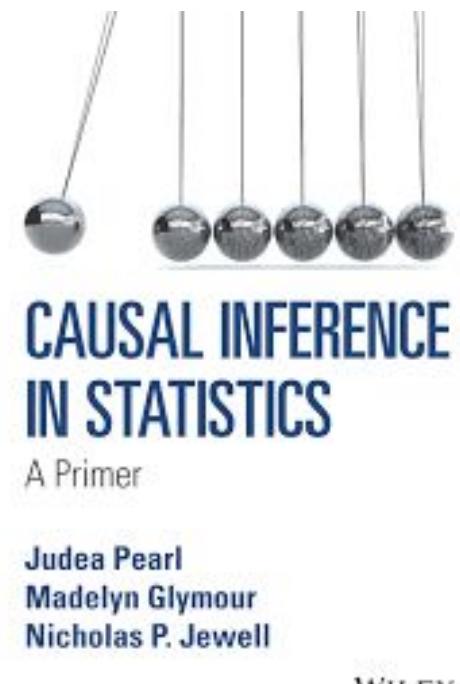
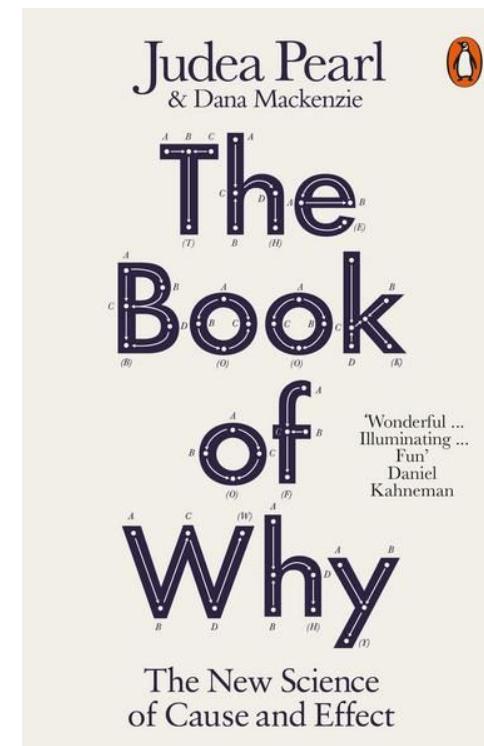


Resources for learning about causal inference

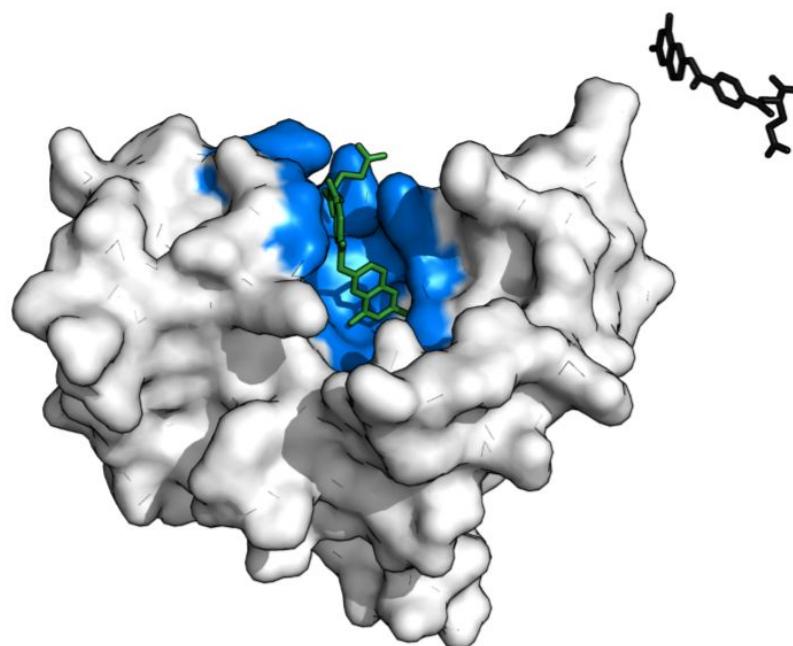
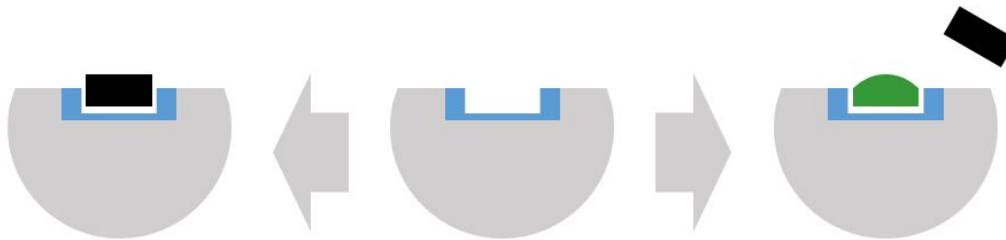
Causal inference in drug discovery and development, Michoel and Zhang, 2022



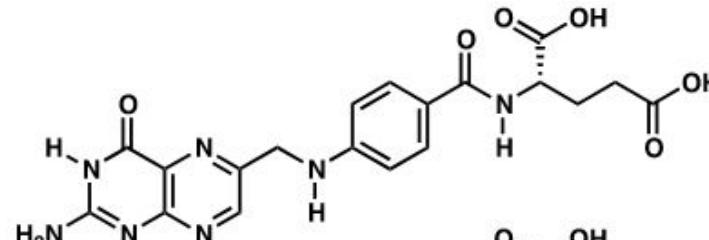
Lectures available on YouTube



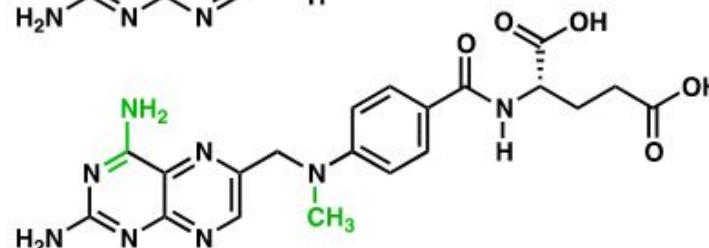
Seeing how a drug works



[Work by Thomas Shafee, Shared under CC-AS-4.0](#),
and work by [Boghog](#). Based on [PDB record 4QI9](#).



Dihydrofolic acid



MTX

The protein: Dihydrofolate reductase (DHFR), which converts dihydrofolic acid into tetrahydrofolate. The process is important for cell proliferation and cell growth. DHFR is a drug target for cancer and autoimmune diseases.

The natural substrate: dihydrofolic acid (vitamin B9), in black. Dihydrofolic acid is the *natural ligand* of DHFR.

The drug: methotrexate (MTX), in green. MTX is a *synthesized ligand* of DHFR, and it is a competitive inhibitor of DHFR with regard to its natural substrate.

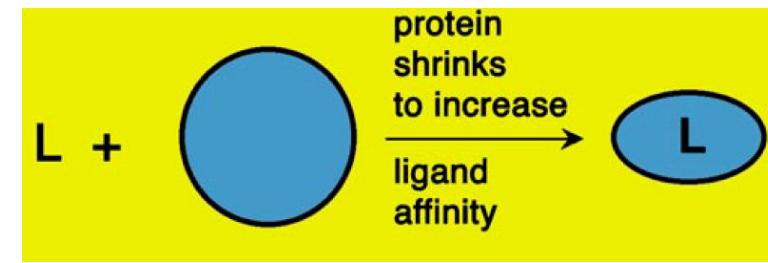
The binding site: where the enzyme binds its substrate and catalyses the chemical reaction, in blue.

Biophysical and biochemical views of ligand-target binding

The biophysical view of binding: Both enthalpy (heat transfer) and entropy (disorder) contribute to the binding energy ($\Delta G = \Delta H - T\Delta S$).

- Binding occurs in favourable steric, *i.e.* spatial, configurations. A simplification is **the ‘lock-and-the-key’ model**, however, in reality enzyme **undergoes changes in its shape**.
- Binding is mediated by intermolecular forces, such as electrostatic interactions (*e.g.* hydrogen bonds), Van der Waals forces (dipole interactions), π -effects (interactions of π -orbitals of a molecular system), and hydrophobic effect.
- Binding opposes motion, and motion opposes binding: there is enthalpy/entropy compensation in ligand-substrate binding.

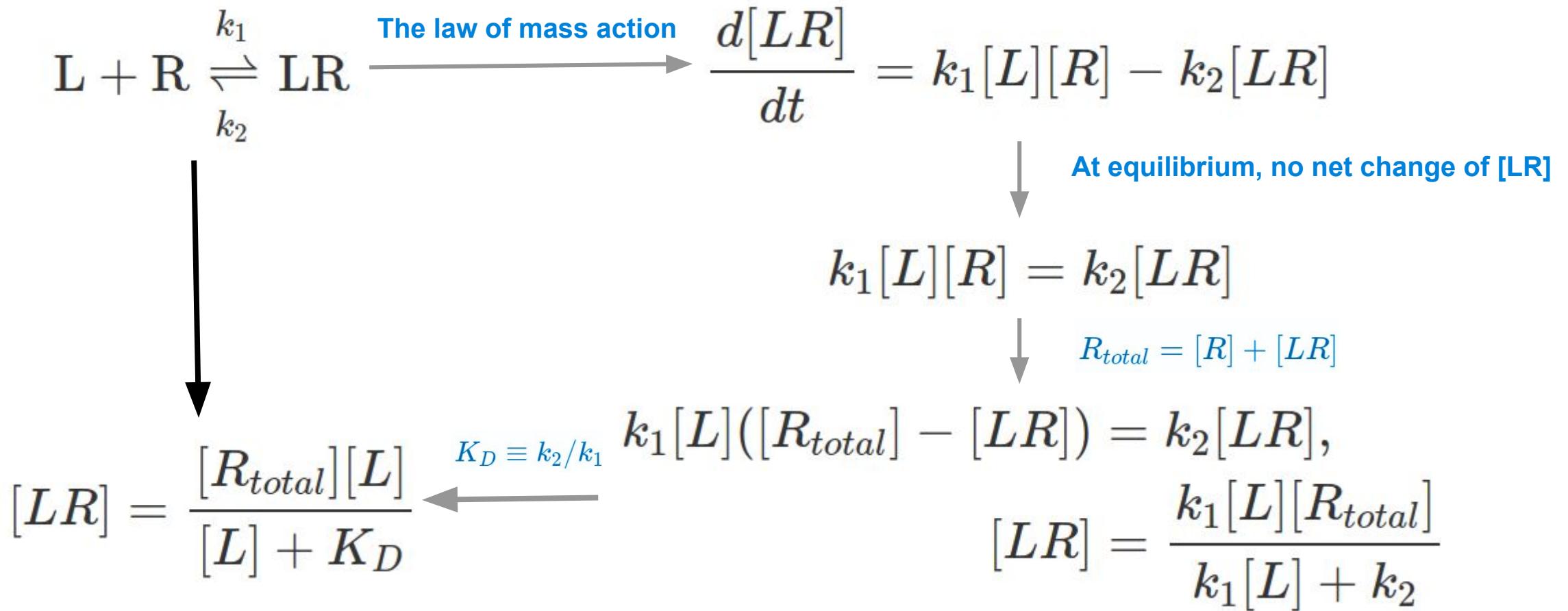
The biochemical view of binding: The *rate* of binding is called affinity, often expressed in K_d or, for inhibitors, K_i . A closely related, and often confusing, concept is IC_{50} .



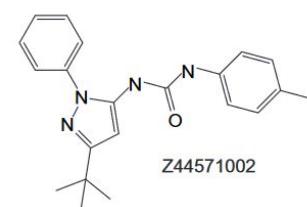
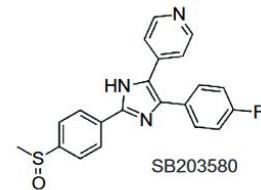
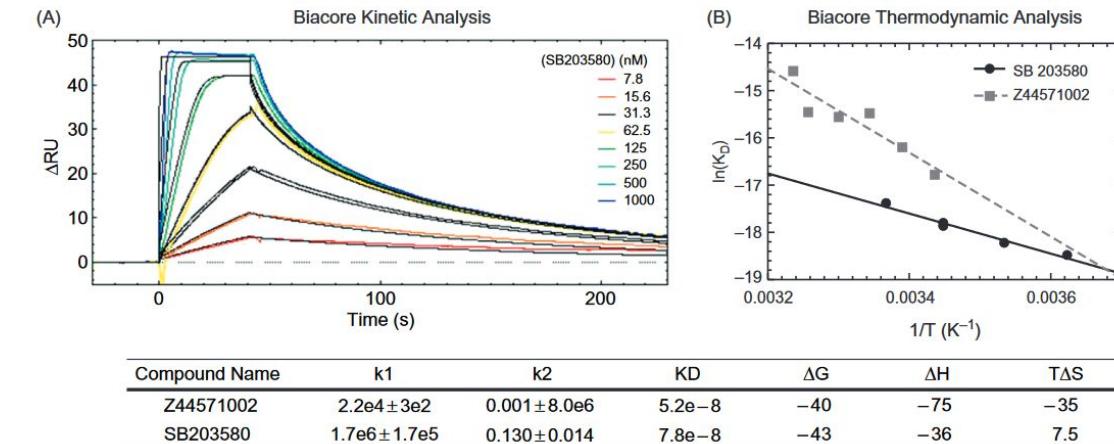
Competitive	Uncompetitive
$E + S \xrightleftharpoons[K_m]{+} ES \xrightarrow{k_{cat}} E + P$ <p style="text-align: center;">$K_i \parallel$ EI</p>	$E + S \xrightleftharpoons[K_m]{+} ES \xrightarrow{k_{cat}} E + P$ <p style="text-align: center;">$K_i \parallel$ ESI</p>
Non-competitive	Mixed
$E + S \xrightleftharpoons[K_m]{+} ES \xrightarrow{k_{cat}} E + P$ <p style="text-align: center;">$K_i \parallel$ $K_i \parallel$ EI + S $\xrightleftharpoons[K_m]{+}$ ESI</p>	$E + S \xrightleftharpoons[K_m]{+} ES \xrightarrow{k_{cat}} E + P$ <p style="text-align: center;">$K_i \parallel$ $\alpha K_i \parallel$ EI + S $\xrightleftharpoons[\alpha K_m]{+}$ ESI</p>

Four basic types of kinetic mechanism of inhibition, source: sciencesnail.com

From the law of mass action to ligand-target interaction



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



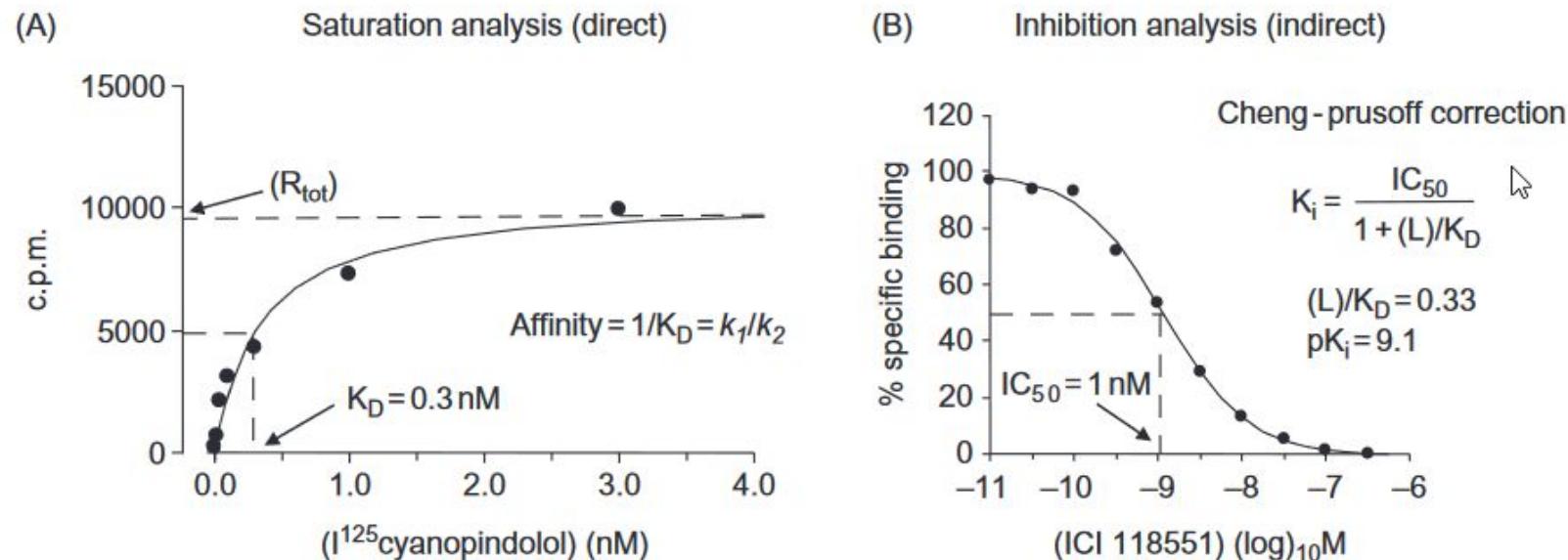
$$\Delta G = \Delta H - T\Delta S \xrightarrow{\Delta G = -RT \ln K} \text{Van't Hoff Equation} \quad -\ln(K_D) = \frac{\Delta H}{R} \frac{1}{T} - \frac{\Delta S}{R}$$

Kinetic and thermodynamic measurements of two p38 α inhibitors.

(A) The time course of SB203580 binding to immobilized mitogen activated kinase p38 α . The y-axis shows the mass change resulting from compound binding to p38 α . At t=0, a range of SB203580 concentrations were passed across the immobilized p38 α 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. ΔG , ΔH 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.

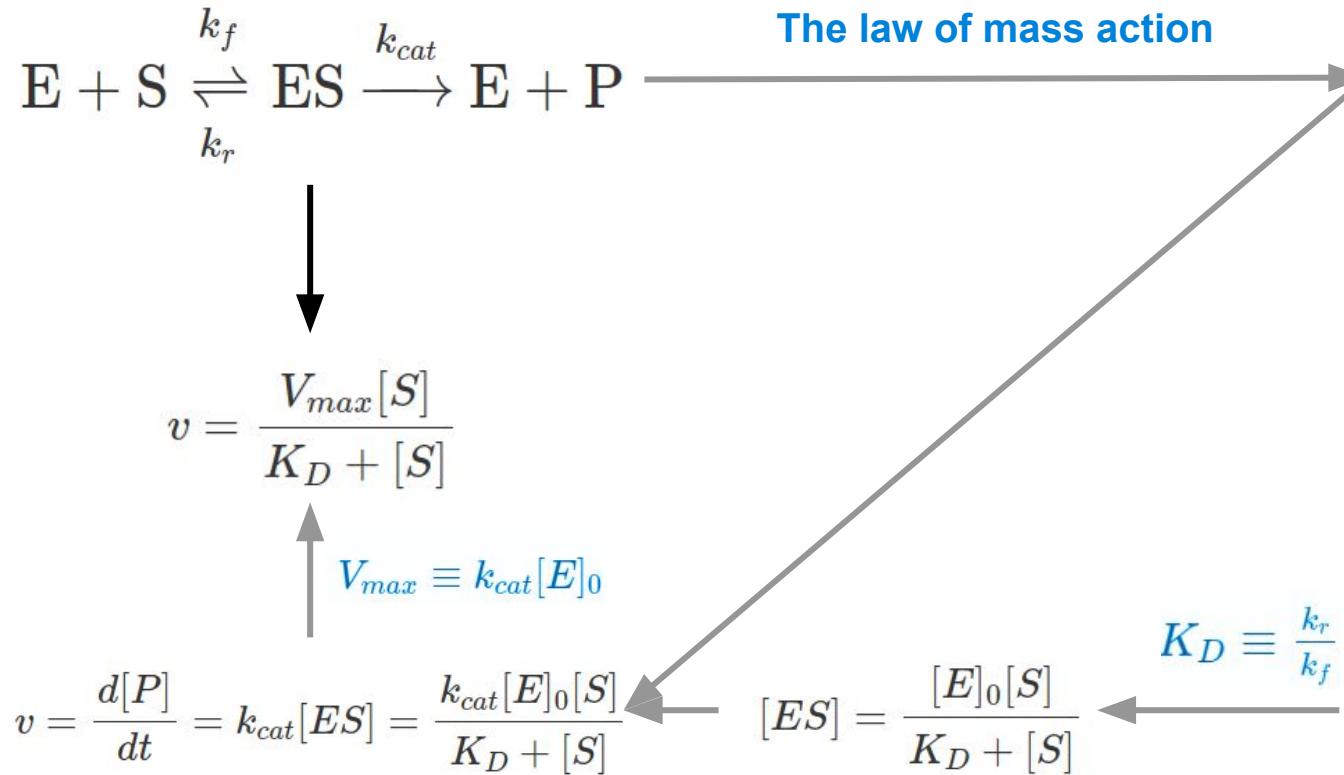
The biochemical (kinetic) view of binding affinity: the hyperbola curve and the dissociation constant K_D



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?

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 IC₅₀: The Hill function and *in vitro* pharmacology

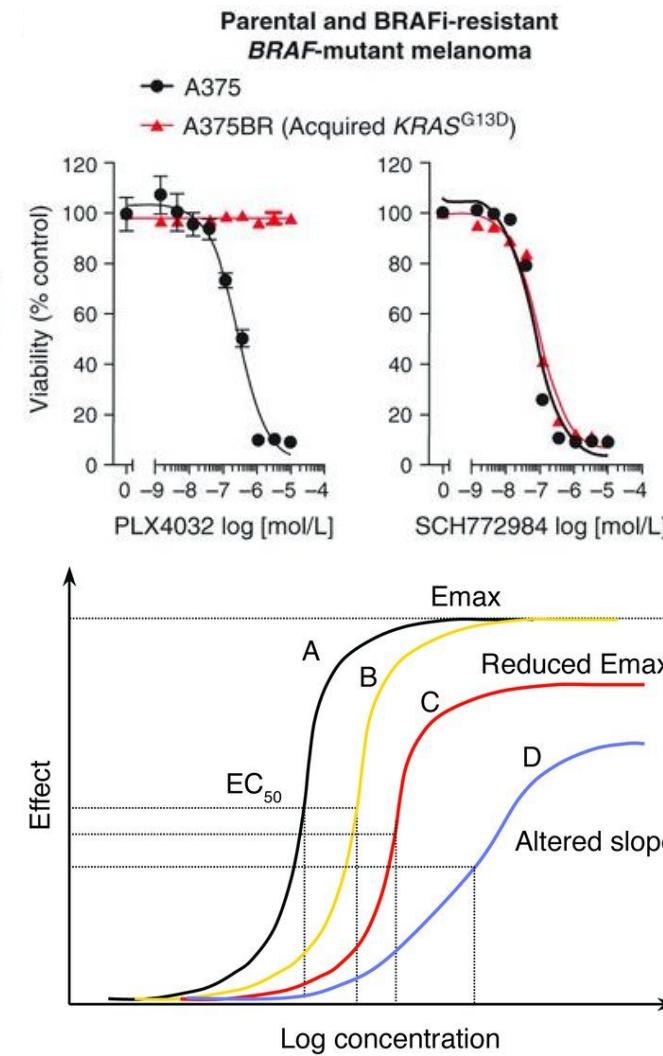
- The Hill function is one of the mostly useful non-linear functions to model biological systems.
- In its general form, H_{max} indicates the maximal value to which the function is asymptotic, n is the shape parameter (known as the Hill's coefficient), and k is the reflection point, often abbreviated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

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

The 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 the dose-dependent effect



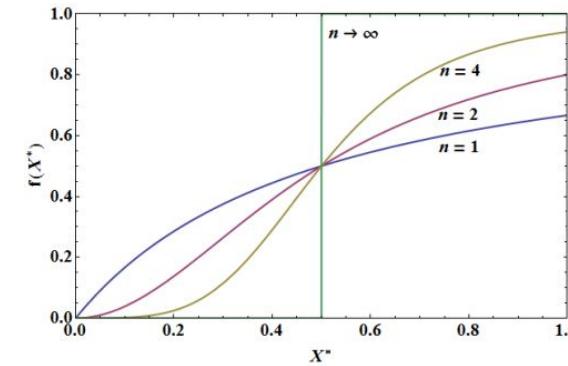
[Morris et al. Cancer Discov; 3\(7\): 742–50.](#)
©2013 AACR.

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

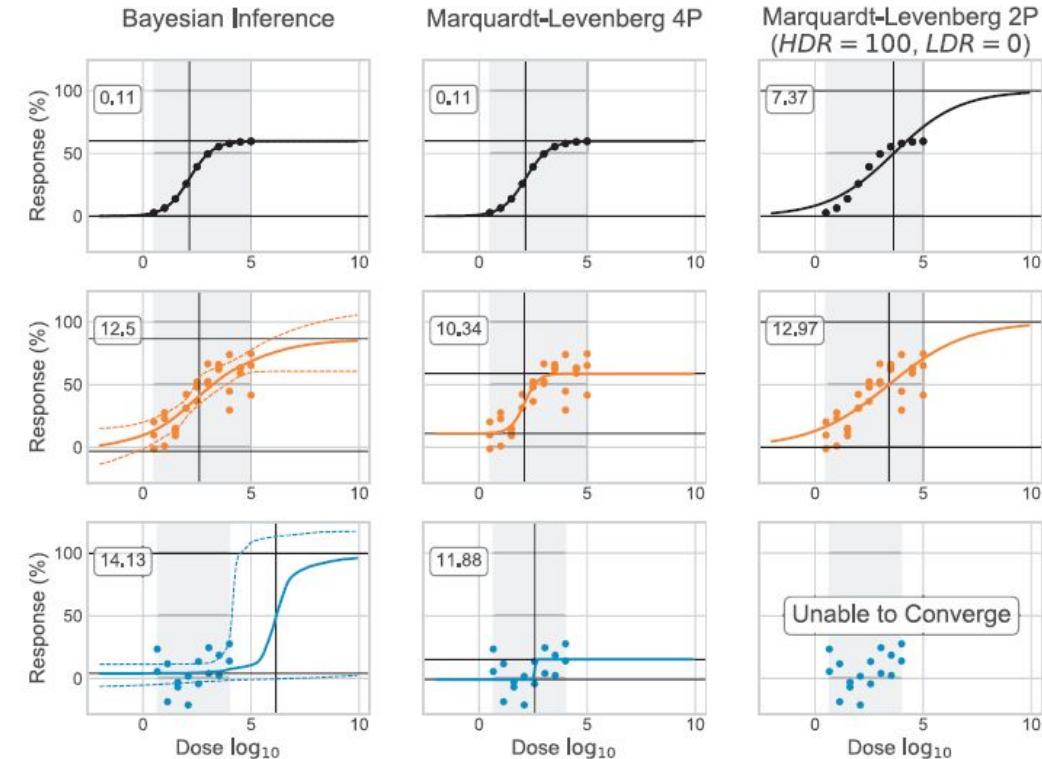
Theoretical and practical considerations about the Hill function

- 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.
- The Hill function is often used to model either *target occupancy* or *tissue response* (pharmacology).
- 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.
- Dose-response data may look quite different from the ideal curve (bottom panel). By using a Bayesian inference approach, it is possible to perform inference even with ill-looking data.

The Bayesian inference approach versus the non-Bayesian Marquardt-Levenberg algorithm for non-linear regression fitting. 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.



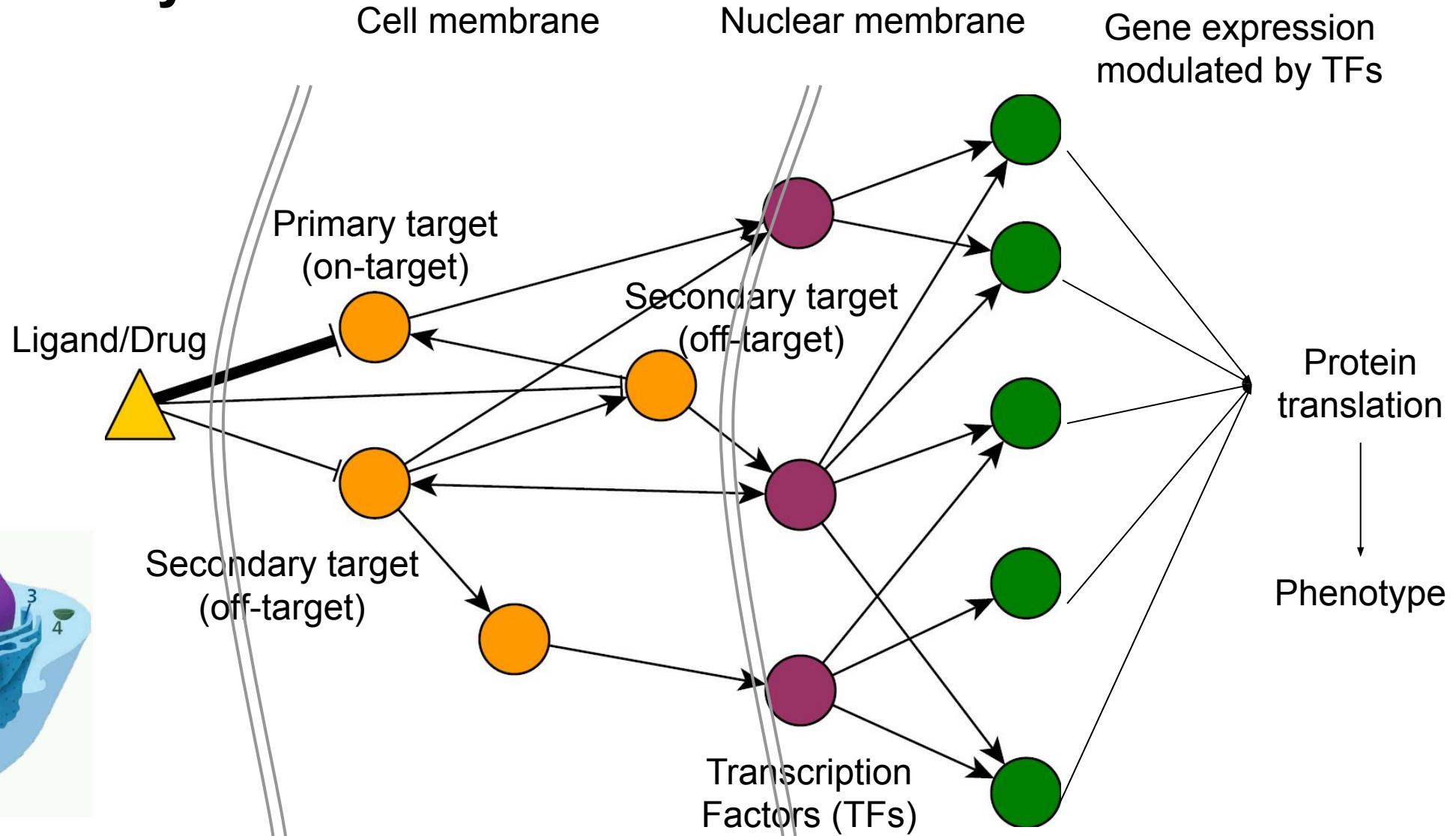
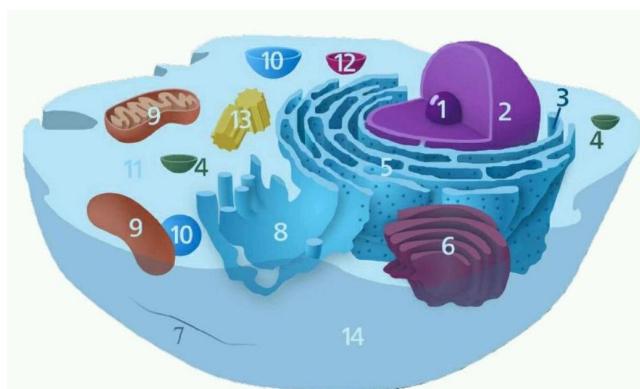
From [the biophysics wiki article](#) by Andreas Piehler



Why modelling molecules is not enough for drug discovery?

The importance of networks

Biological networks interact with drugs and manifest its efficacy and safety



Summary

- Machine learning is important in drug discovery. Machine learning should be guided by chemical and biological models to improve human understanding.
- ODE-based compartment models and stochastic models can be used to model small to moderate biochemical networks.
- The network nature of biology requires models beyond the molecular level.

Five classes of mathematical models drug discovery

Compartment models

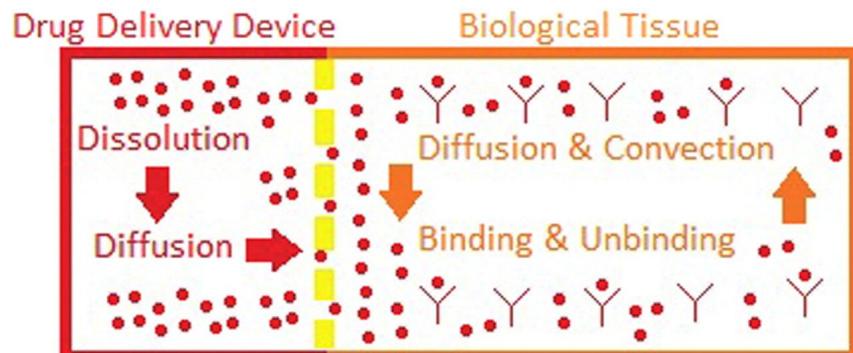
$$\begin{aligned}\frac{dx}{dt} &= \alpha x - \beta xy, \\ \frac{d[LR]}{dt} &= k_1[L][R] - k_2[LR] \\ \frac{dy}{dt} &= -\gamma y + \delta xy,\end{aligned}$$

Kinetics of ligand-target interaction

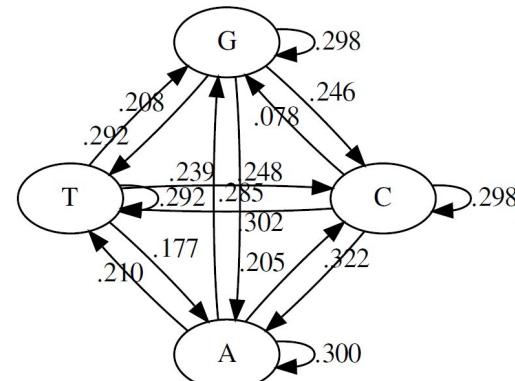
$$\begin{aligned}\frac{dS}{dt} &= -\frac{\beta IS}{N}, \\ \frac{dI}{dt} &= \frac{\beta IS}{N} - \gamma I, \\ \frac{dR}{dt} &= \gamma I\end{aligned}$$

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

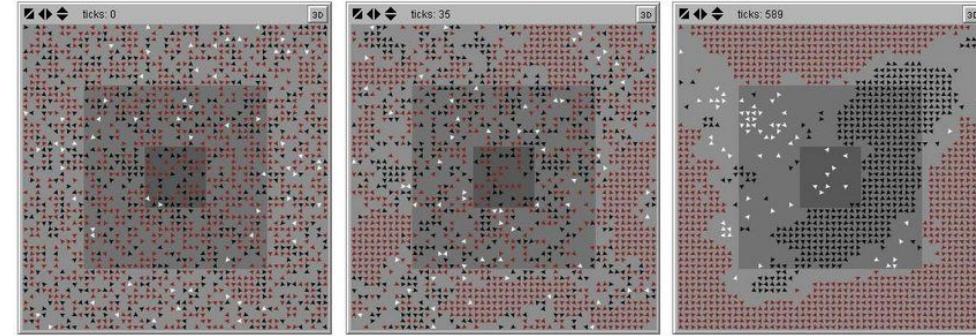
Transport models



Finite state models

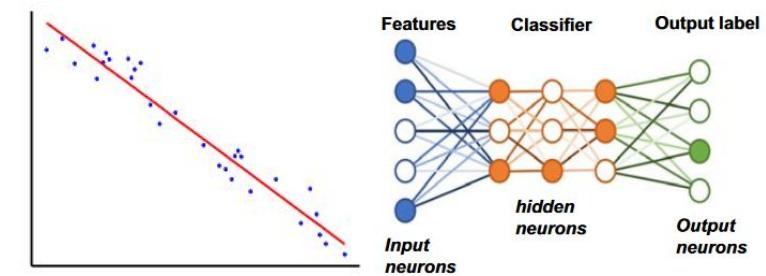


Particle models



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

Statistical/machine learning models



Backup slides

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**: the basic idea is that 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**, for instance using Coulombic interactions and L-J 12-6 function.

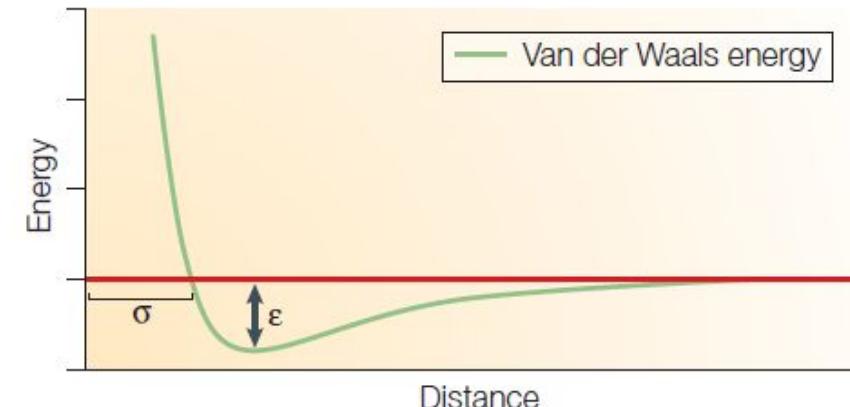
$$E_{coul}(r) = \sum_{i=1}^{N_A} \sum_{j=1}^{N_B} \frac{q_i q_j}{4\pi\epsilon_0 r_{ij}}$$

Coulombic interactions
(electrostatic interactions between electric charges)

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

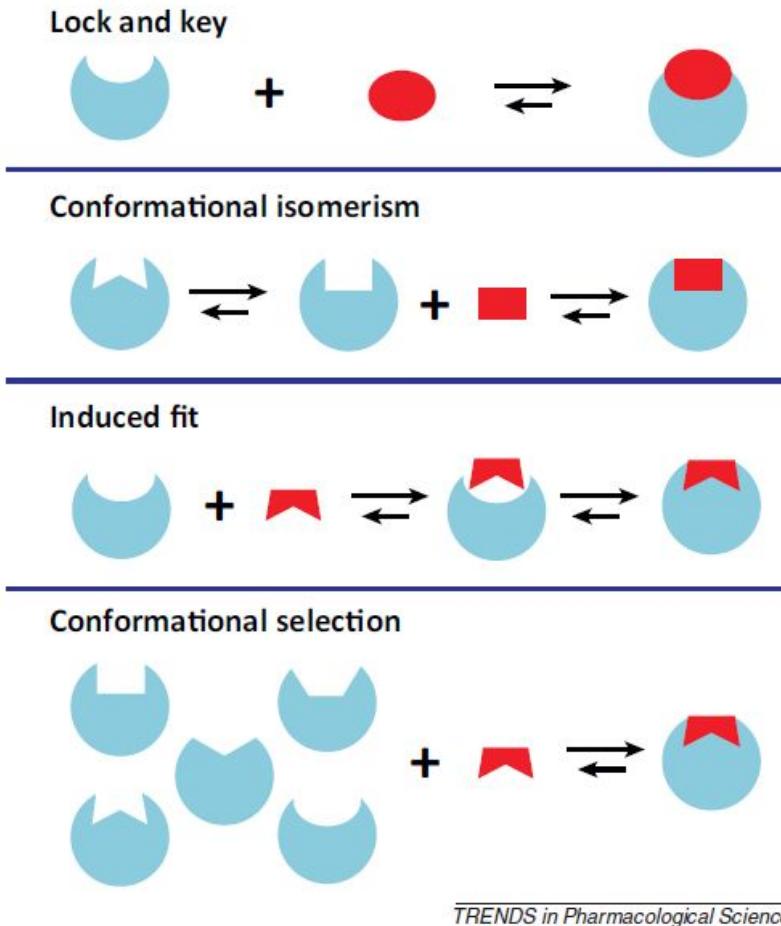
Lennard–Jones 12–6 function
(intermolecular interactions without charge)

- ϵ is the **well depth** of the potential
- σ is the **collision diameter** of the respective atoms i and j .

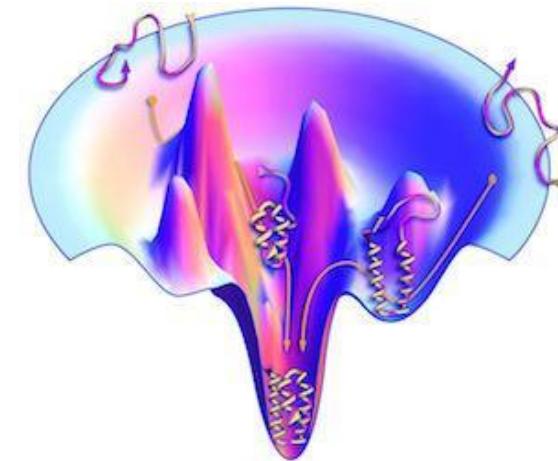
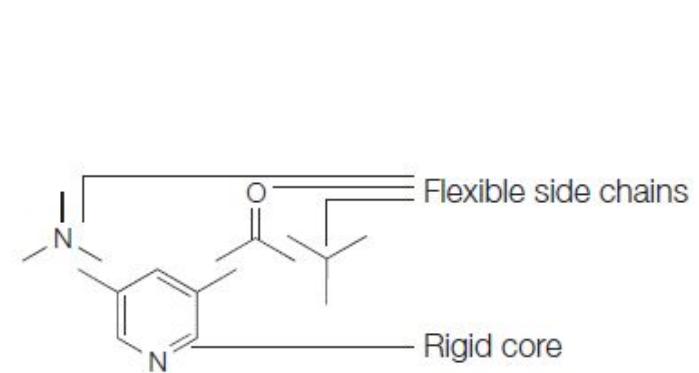


Kitchen, Douglas B., Hélène Decornez, John R. Furr, und Jürgen Bajorath. „Docking and Scoring in Virtual Screening for Drug Discovery: Methods and Applications“. *Nature Reviews Drug Discovery* 3, Nr. 11 (November 2004): 935–49. <https://doi.org/10.1038/nrd1549>.

Posing: dealing with flexibility of ligand and of protein



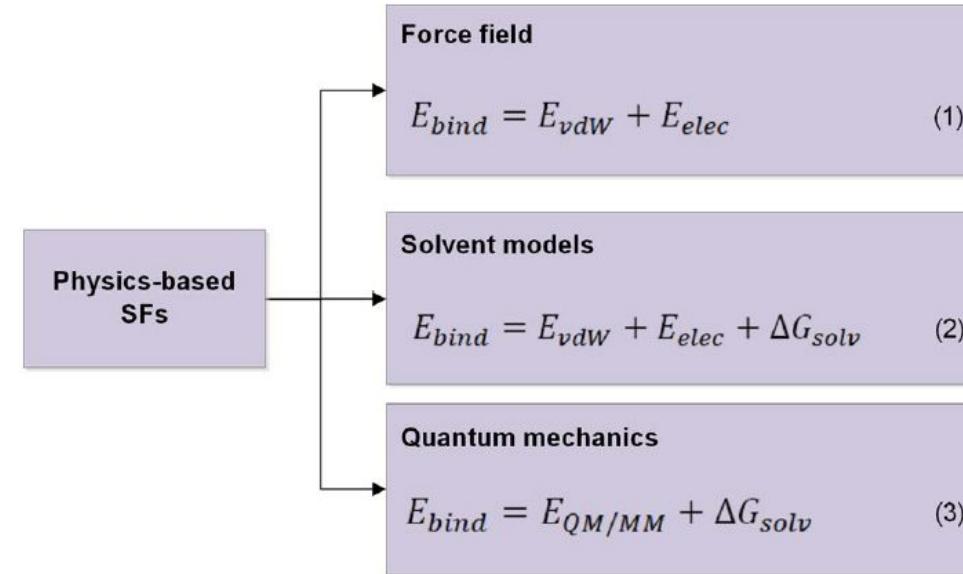
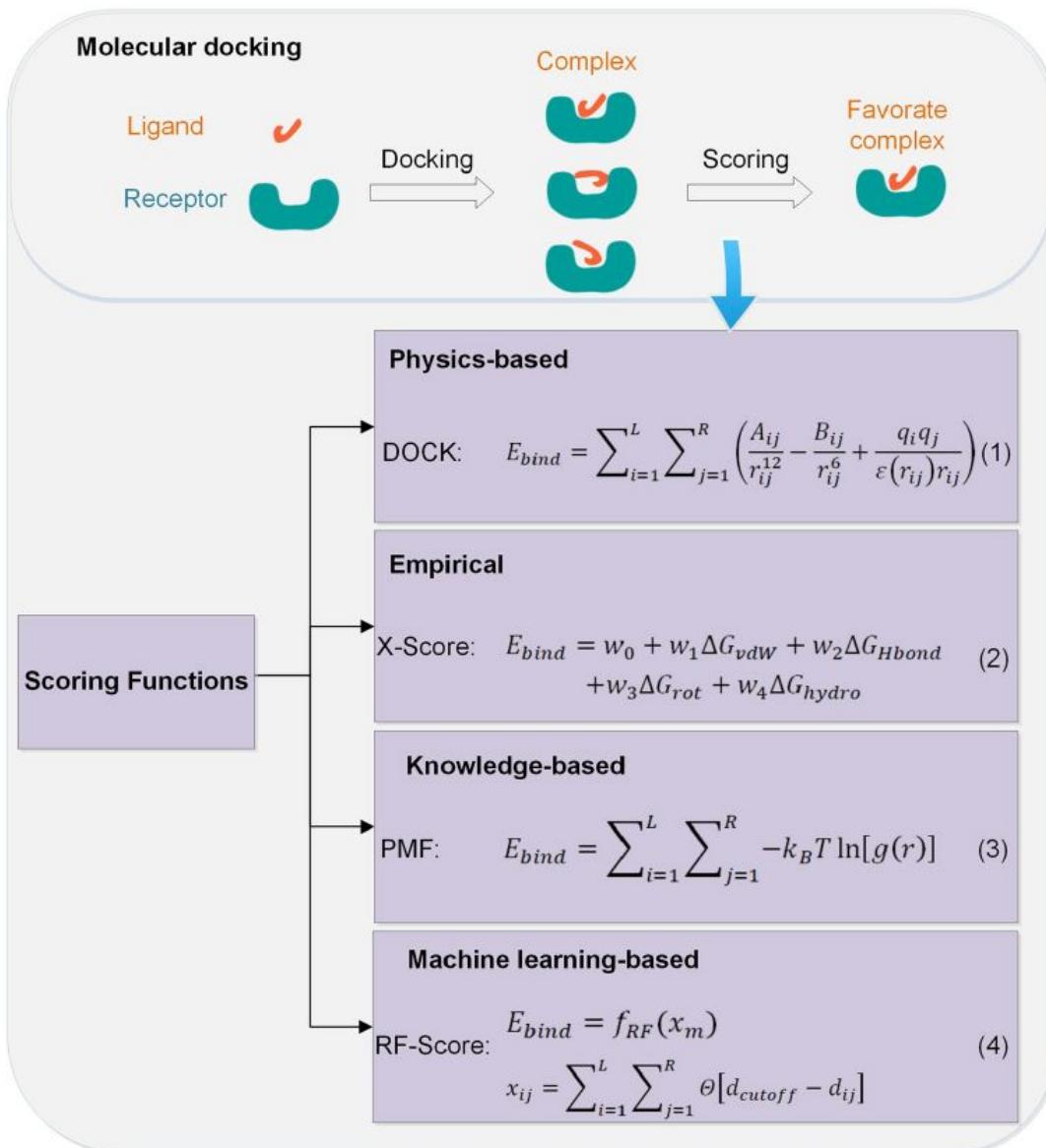
Chen, Yu-Chian. „Beware of docking!“ *Trends in Pharmacological Sciences* 36, Nr. 2 (1. Februar 2015): 78–95.
<https://doi.org/10.1016/j.tips.2014.12.001>.



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](https://doi.org/10.1038/nprot.2016.051)

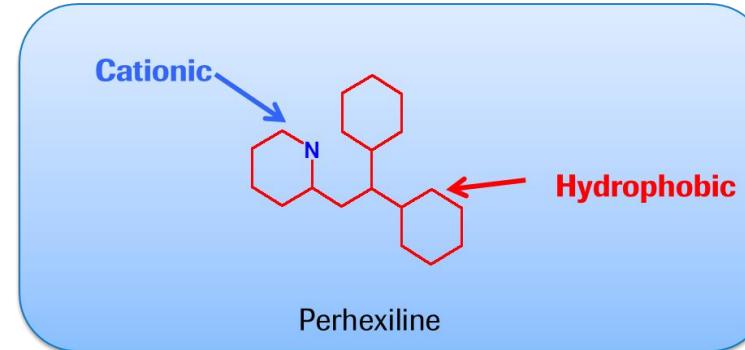
Computational docking can be used to predict bound conformations and free energies of binding for small-molecule ligands to macromolecular targets. Docking is widely used for the study of biomolecular interactions and mechanisms, and it is applied to structure-based drug design. The methods are fast 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.
- **Binding predicted by docking should always be challenged and verified by experimental testing! Docking scores seldomly correlate with binding affinity.**

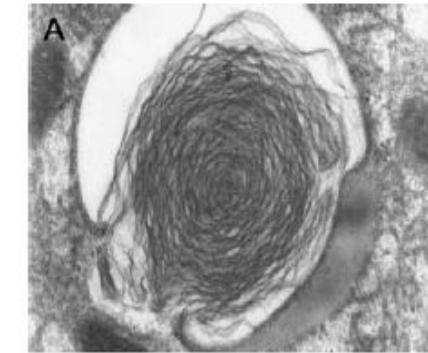
Drug-induced phospholipidosis is correlated with amphiphilicity

- 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

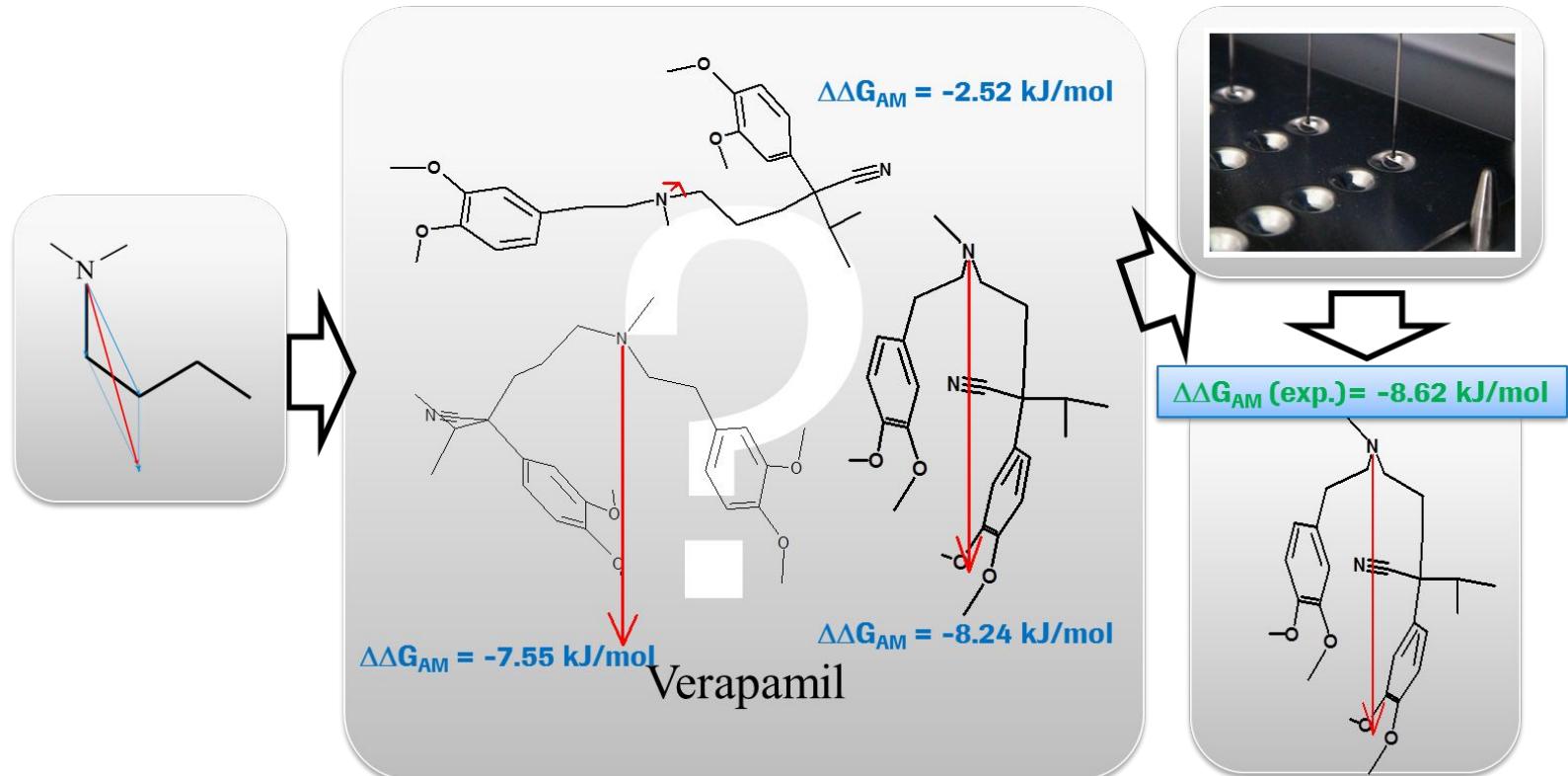
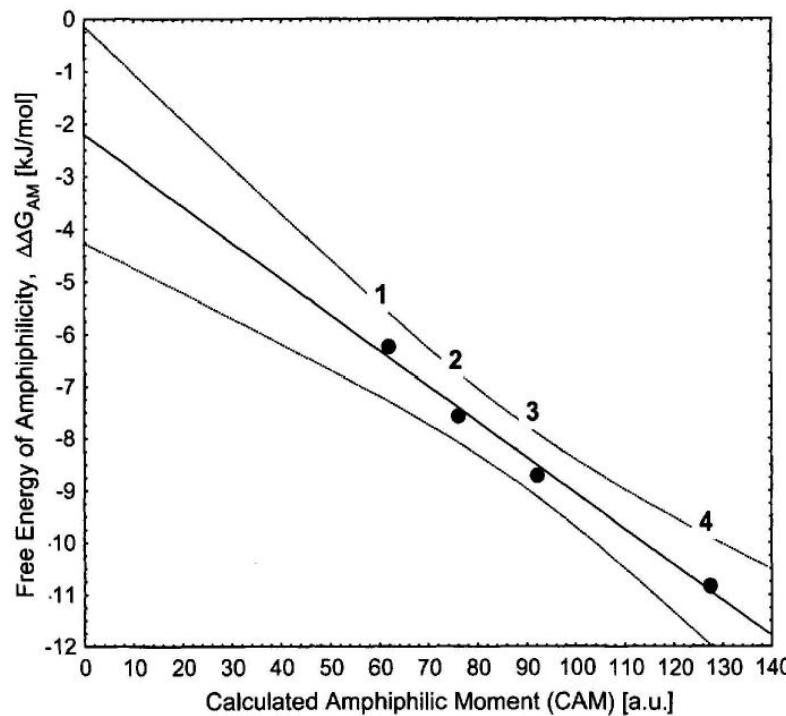
d : distance between the center of gravity of the charged part of a molecule and the hydrophobic/hydrophilic remnant of the molecule

$\vec{\alpha}_i$: 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

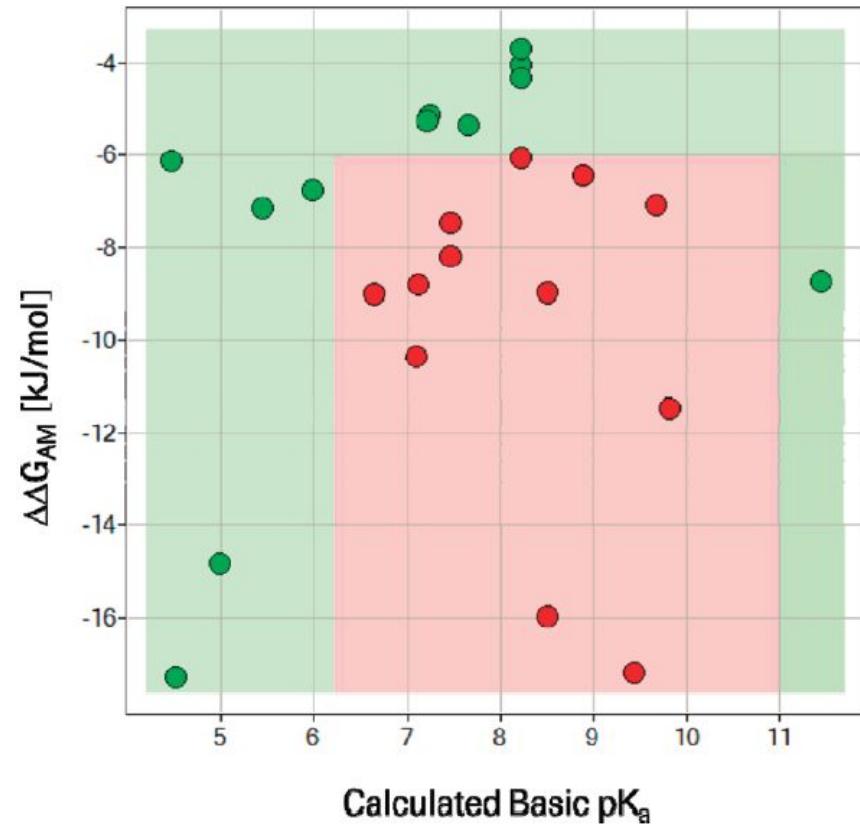
Development of CAFCA (CAlculated Free energy of amphiphilicity of small Charged Amphiphiles)



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 phospholipidosis 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 (and lessons not yet 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).
- Safety liabilities caused by physicochemical properties of the drugs may be well predicted by molecular modelling inspired by simple models.**

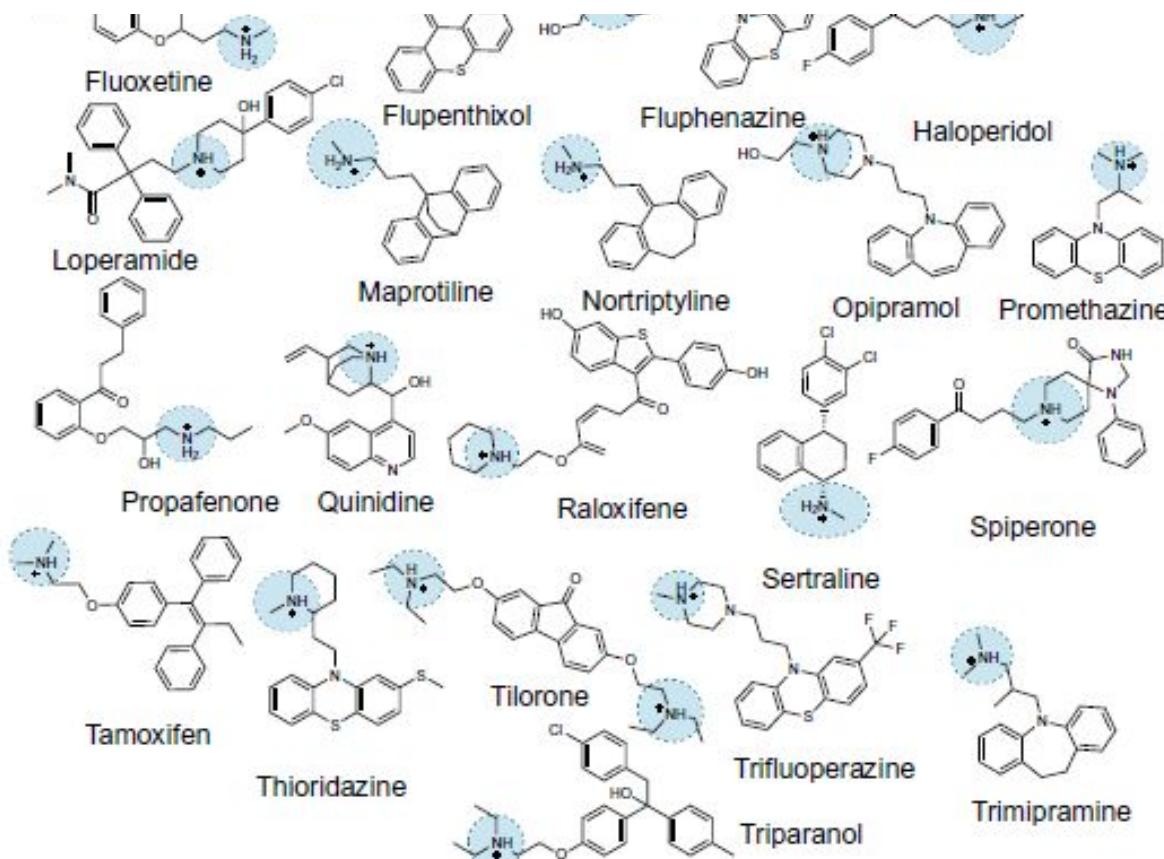


Fig. 1. Representative examples of CADs that are identified in SARS-CoV-2 drug repurposing screens.

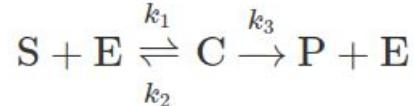
Tummino, Tia A., Veronica V. Rezelj, Benoit Fischer, Audrey Fischer, Matthew J. O'Meara, Blandine Monel, Thomas Vallet, et al. "Drug-Induced Phospholipidosis Confounds Drug Repurposing for SARS-CoV-2." *Science* 373, no. 6554 (July 30, 2021): 541–47. <https://doi.org/10.1126/science.abi4708>.

More about the 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)

Simulation of biological networks with ordinary differential expression: the simplest case

Given the reaction

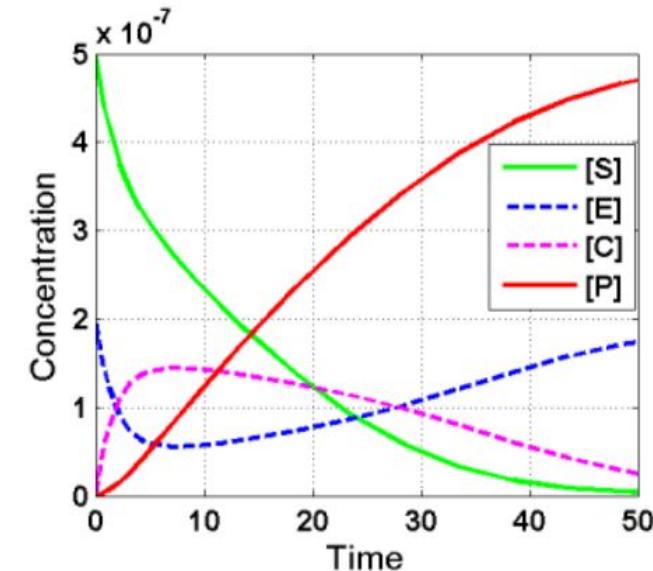


According to the law of mass action

$$\begin{aligned}\frac{d[S]}{dt} &= -k_1[E][S] + k_2[C], \\ \frac{d[E]}{dt} &= -k_1[E][S] + (k_2 + k_3)[C], \\ \frac{d[C]}{dt} &= k_1[E][S] - (k_2 + k_3)[C], \\ \frac{d[P]}{dt} &= k_3[C],\end{aligned}$$

See [Systems Engineering Wiki \(tue.nl\)](#) for MATLAB/COPASI codes and *Stochastic Modelling for Systems Biology* by Darren J. Wilkinson

- Given the initial values and rate constants
- $S(0) = 5e^{-7}$
 - $E(0) = 2e^{-7}$
 - $C(0) = P(0) = 0$
 - $k_1 = 1e^6$
 - $k_2 = 1e^{-4}$
 - $k_3 = 0.1$



It is possible to simulate the concentration changes by time *deterministically*.

Chemical Master Equations (CME): a particle model of chemical reaction

Given the reaction $A + B \xrightleftharpoons[k_2]{k_1} C + D$ and the initial condition $X(0) = \begin{bmatrix} K \\ K \\ 0 \\ 0 \end{bmatrix}$ (K molecules of species A and of species B respectively)

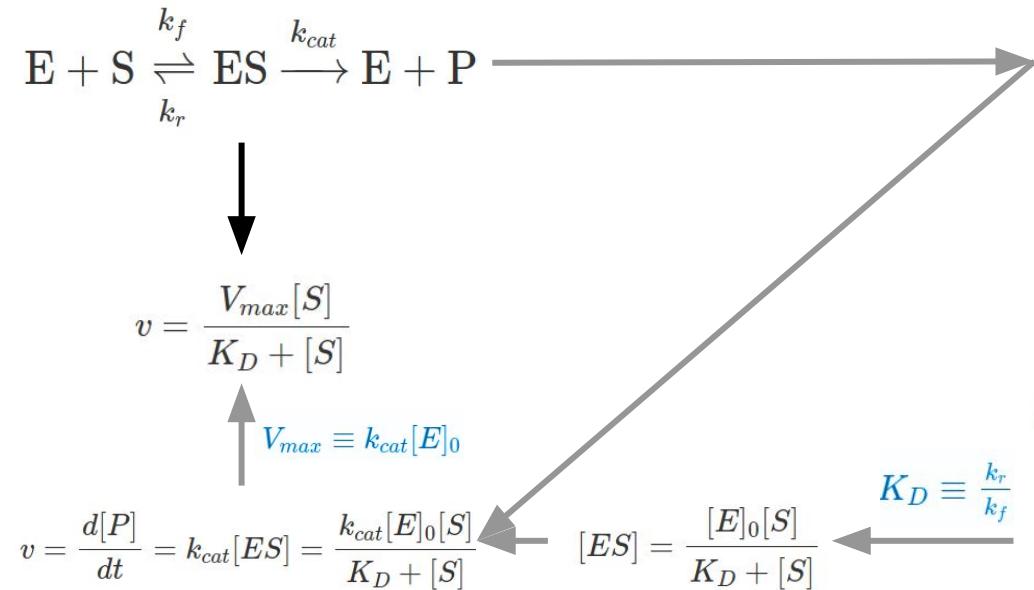
The state vector $X(t)$ can take at any time point *one* of the values

$$\begin{bmatrix} K \\ K \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} K-1 \\ K-1 \\ 1 \\ 1 \end{bmatrix}, \begin{bmatrix} K-2 \\ K-2 \\ 2 \\ 2 \end{bmatrix}, \dots, \begin{bmatrix} 0 \\ 0 \\ K \\ K \end{bmatrix},$$

Theoretically we can build an ODE system with $K+1$ equations to model *every state of the reaction*, down to every particle. In reality, the dimension is so high so that a simulation is not feasible.

CME is a set of ODEs, with each ODE representing one possible state of the system. Solution of the k th equation at time t is a real number giving the probability of system being in that particular state at that time.

Reaction Rate Equations (RRE): a compartment model

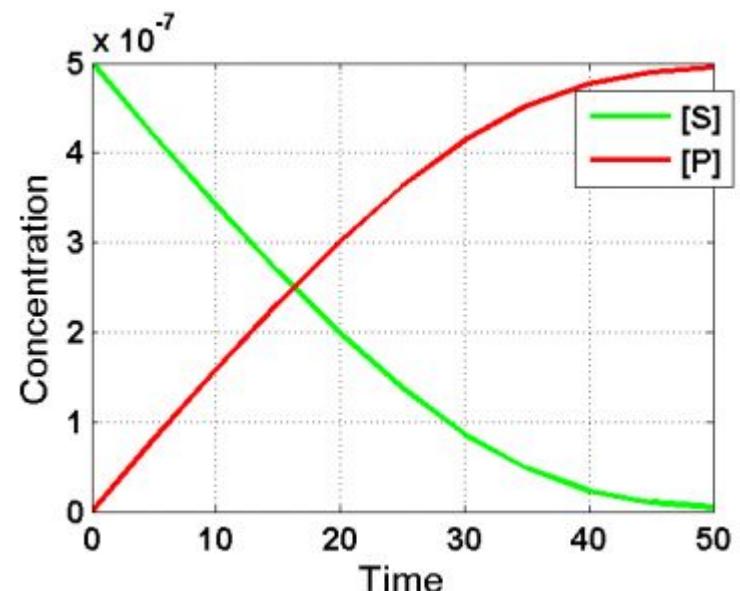


$$\begin{aligned} \frac{d[E]}{dt} &= -k_f[E][S] + k_r[ES] + k_{cat}[ES], \\ \frac{d[S]}{dt} &= -k_f[E][S] + k_r[ES], \\ \frac{d[ES]}{dt} &= k_f[E][S] - k_r[ES] - k_{cat}[ES], \\ \frac{d[P]}{dt} &= k_{cat}[ES], \end{aligned}$$

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

$$\begin{aligned} k_f([E]_0 - [ES])[S] &= k_r[ES] \\ k_f[E]_0[S] - k_f[ES][S] &= k_r[ES] \\ k_f[E]_0[S] &= k_r[ES] + k_f[ES][S] \\ k_f[E]_0[S] &= [ES](k_r + k_f[S]) \\ [ES] &= \frac{k_f[E]_0[S]}{k_r + k_f[S]} \\ [ES] &= \frac{k_f[E]_0[S]}{k_f(\frac{k_r}{k_f} + [S])} \end{aligned}$$

RRE simulation of the Michaelis-Menten model



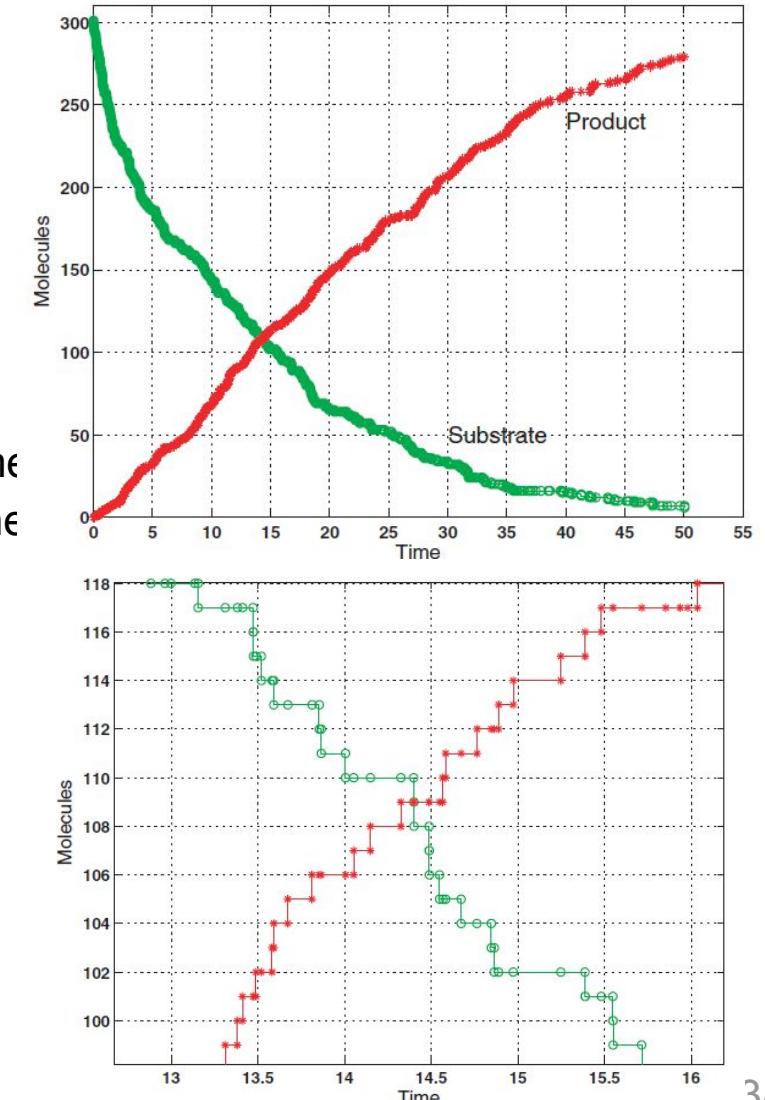
Source: [Systems Engineering Wiki \(tue.nl\)](#)

RRE is a set of ODEs, with each ODE representing one chemical species. Solution of the j th equation at time t is a real number representing the concentration of species j at time t .

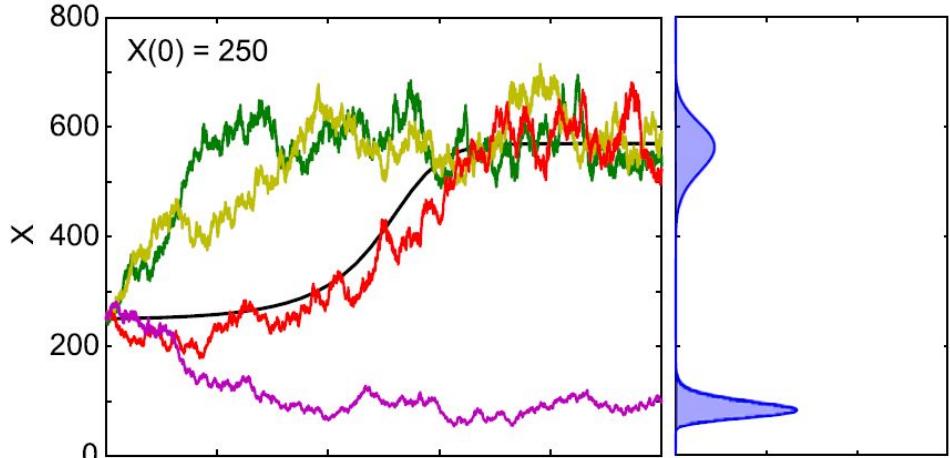
The Gillespie's algorithm and the chemical Langevin equation allow stochastic simulation of biological networks

- The *stochastic simulation algorithm* (exact SSA), also called *Gillespie's algorithm*, allows stochastic simulation of a reaction. It is done in four steps:
 - initialize** the system with initial conditions
 - Given a state at time t , we can define a probability p that reaction j takes place in the time interval $[t+\tau, t+\tau+d\tau]$. It is the product of two density functions of two random variables: the probability of reaction j happens (proportional to the number of substrate molecules), multiplied by the time until next reaction, which is exponentially distributed. This is known as the **Monte Carlo** step.
 - Let the randomly selected reaction happen and **update** the time.
 - Iterate** until substrates are exhausted or simulation time is over.
- Further computation tricks, .e. ‘tau-leaping’, are used to lump together reactions. The chemical Langevin equation (CLE) further accelerates stochastic simulation by approximating *Poisson* with normal distribution.

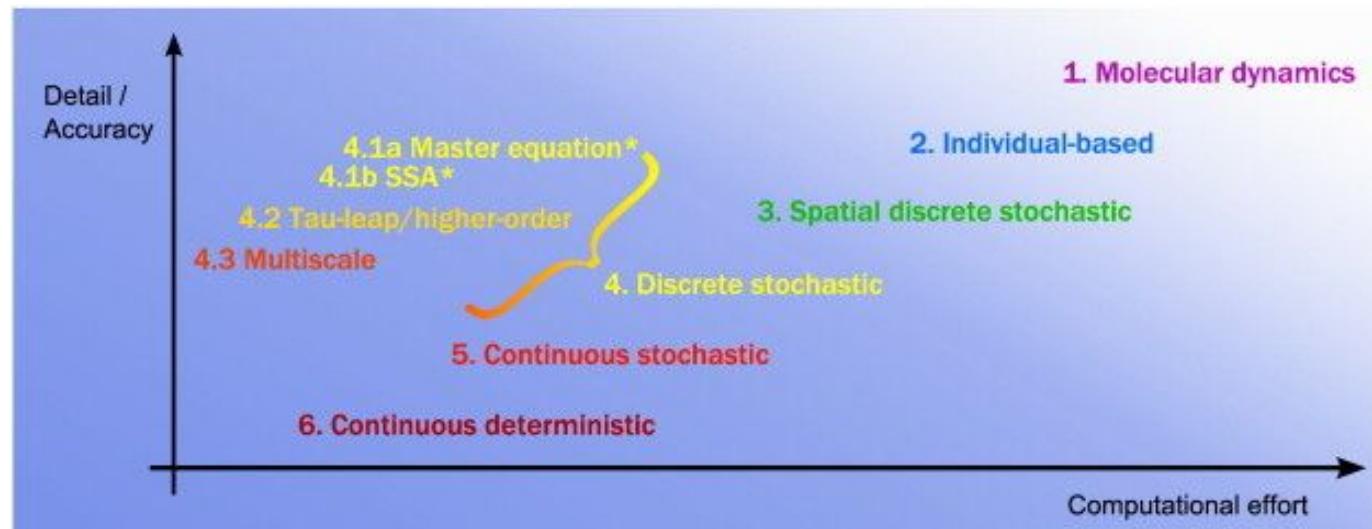
Figure source and further reading: Higham, Desmond J. 2008. “Modeling and Simulating Chemical Reactions.” *SIAM Review* 50 (2): 347–68. <https://doi.org/10.1137/060666457>.



Why stochastic modelling?



- Stochastic modelling can reveal individual trajectories that are otherwise ‘averaged’ by ODE models.
- Small systems and single-molecule studies show stochastic behaviour.
- It is possible to consider both extrinsic and intrinsic factors and take them into the model.



Advantages and disadvantages of several modelling/simulation methods.

Simulation method	Cat.	Advantages	Disadvantages	References	Software
Master equation	4	Exact	Very computationally intensive	[85,143]	
SSA	4	Statistically exact	Very computationally intensive	[82,109]	COPASI [144] StochKit [145] STOCKS [146] BioNetS [147] StochKit [145]
Tau-leap	4	Relatively fast	Approximate; too slow for large systems or frequent/multiscale reactions	[83,113,118]	
Higher-order	4	Relatively fast; accurate	Approximate; too slow for large systems or frequent/multiscale reactions	[83,121,122,124,125]	
Multiscale/hybrid	4	Fast; good for systems with disparate reaction scales	Approximate; problems with coupling different scales	[131,132,137,139,148]	COPASI [144] BioNetS [147]
Brownian dynamics	2	Tracks individual molecules	Slow; molecule size must be artificially added	[149,150]	Smoldyn [149,151] MCell [152]
Compartment-based	3	Accounts for diffusion between homogeneous compartments	Slow; compartment size must be set manually; each compartment is homogeneous	[150,153,154]	MesORD [153] URDME [155]
SDE	5	Fast	Continuous; Gaussian noise	[76]	BioNetS [147]
PDE (R-D)	6	Very fast; spatial	Continuous; no noise	[156]	
ODE	6	Very fast	Continuous; no noise	[157]	

Székely and Burrage. 2014. “[Stochastic Simulation in Systems Biology](#).”

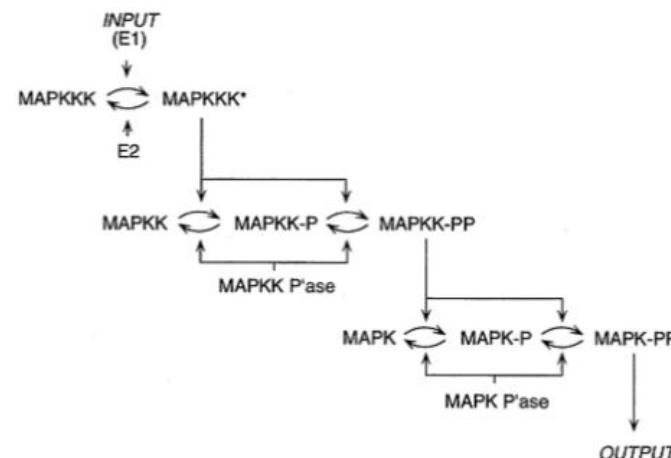
Computational and Structural Biotechnology Journal 12 (20–21): 14–25.

Also see *Stochastic Modelling for Systems Biology* by Darren J. Wilkinson.

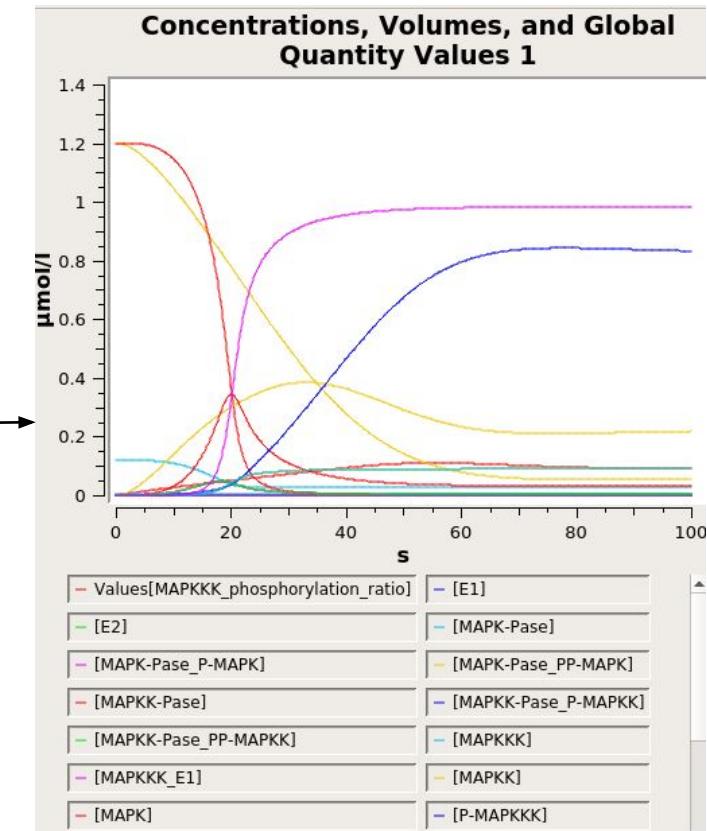
Cat. represents Category from Fig. 2. Abbreviations: SSA, stochastic simulation algorithm; SDE, stochastic differential equation; PDE (R-D), partial differential equation (classical reaction-diffusion equations); ODE, ordinary differential equation.

Biochemical system simulator COPASI

- Freely available at <http://COPASI.org/>
- COPASI supports two types of simulations:
 (1) **ordinary differential equation (ODE)**
 based simulation, (2) **stochastic kinetic simulation**, among others using the stochastic Runge–Kutta method (RI5) and Gillespie's algorithm
 - Resources to learn more about stochastic modelling: [MIT OpenCourseWare](#) by Jeff Gore, and [Stochastic Processes: An Introduction, Third Edition](#) by Jones and Smith
- Tutorials also available on [the website of European Bioinformatics Institute \(EBI\)](#)
- The mathematical concept and software tools are important for detailed analysis of enzymatic reactions, especially in the presence of drugs and/or disease-relevant mutation

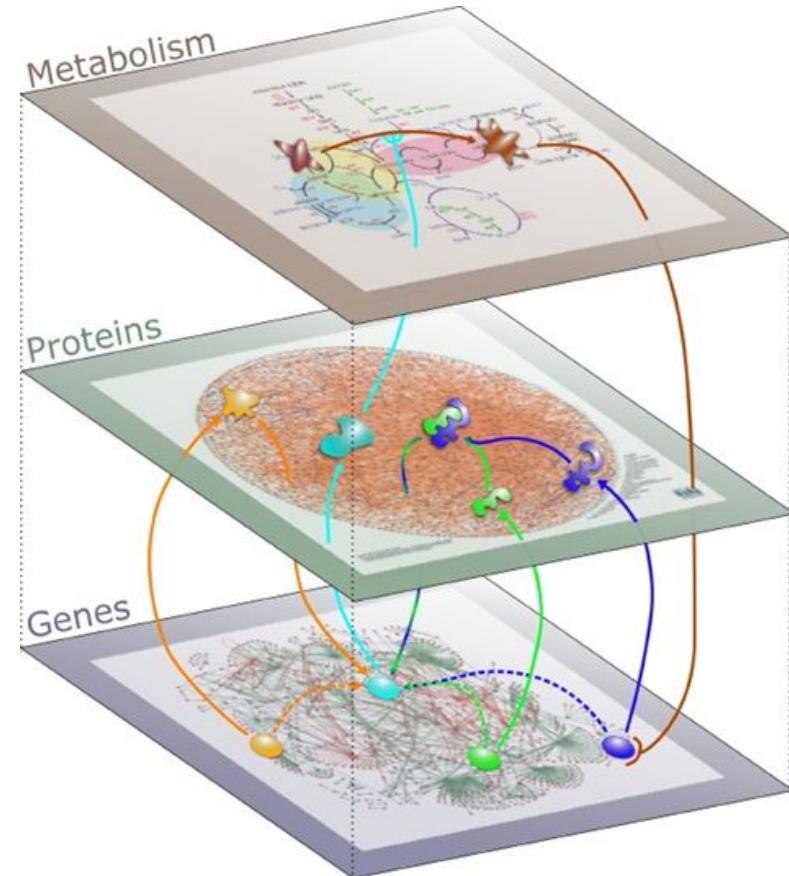


Huang and Ferrell, PNAS, 2006

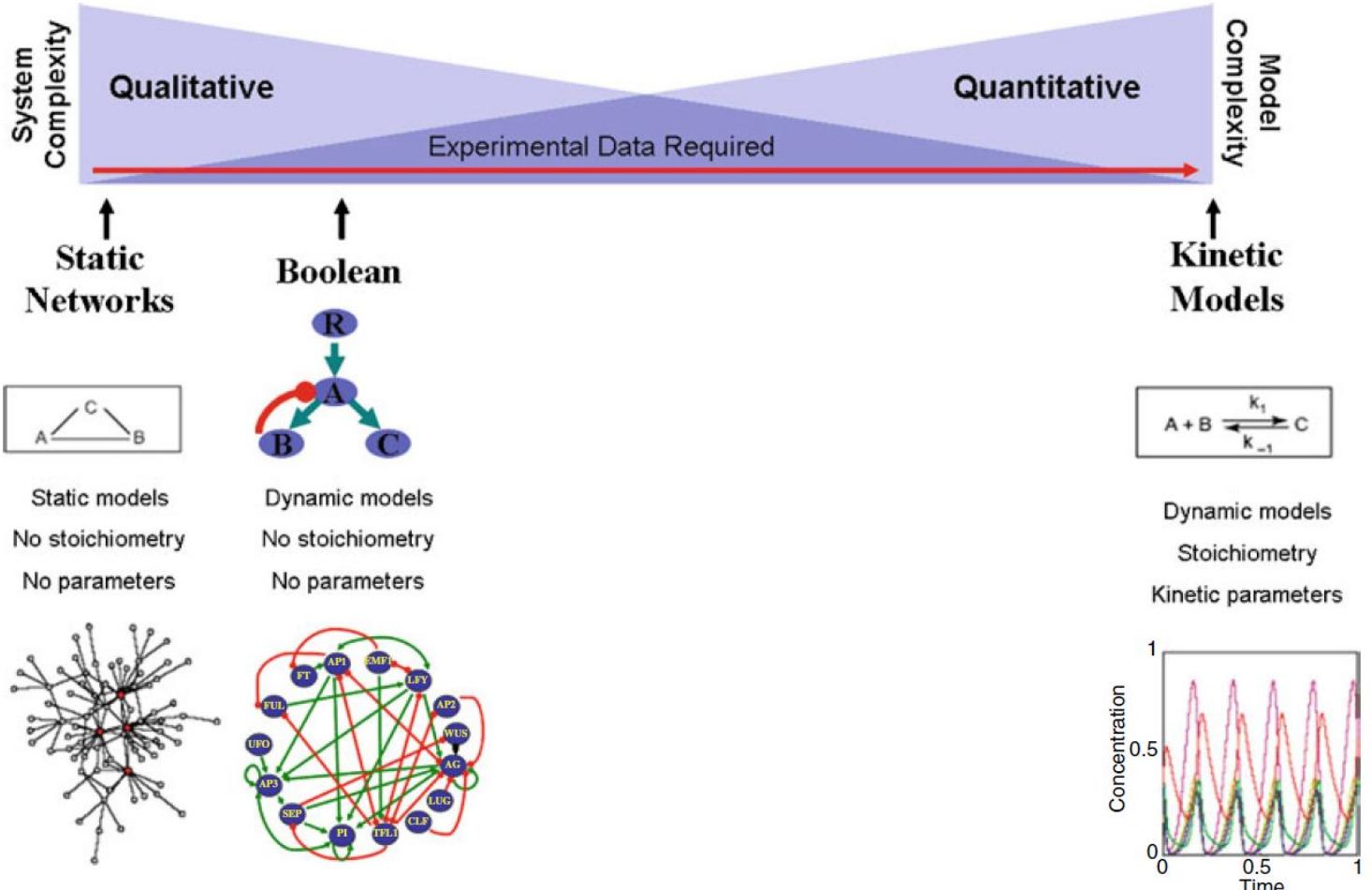


ODE-based simulation of dynamics

Modelling biological networks



Stéphane CHÉDIN & Jean LABARRE, www-dsv.cea.fr



Garg, Abhishek, Kartik Mohanram, Giovanni De Micheli, and Ioannis Xenarios. 2012. "[Implicit Methods for Qualitative Modeling of Gene Regulatory Networks](#)." In *Gene Regulatory Networks: Methods and Protocols*, edited by Bart Deplancke and Nele Gheldof, 397–443. Methods in Molecular Biology. Totowa, NJ: Humana Press.