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? (An ODE model of reaction rate and reactant mass)

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})? (K_D)

Q3: In Figure 2.3, what do x- and y-axis represent in panel (A) and panel (B), respectively? (concentrations in in x-axis; y-axis: counts per minute of radioactivity (A), percentage of binding of the labelled compound)

Q4: What is a sigmoidal curve? (A S-shaped, logistic or logit curve)

Q5: Do IC_{50} values indicate a particular mechanism of action (MoA)? (No)

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? (B>A)

Q7: Why is DMSO often used in bioassays? (solvent, control)

Q8: Can you use your own language to describe what is the Hill function? (discussed in this lecture)

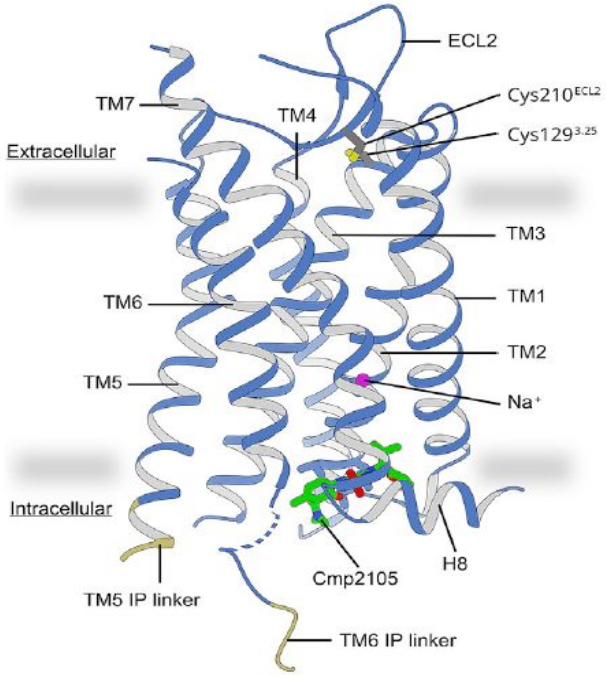
Q9: What statistical measure is used to measure the signal-noise ratio in screening? Can you use your own language explaining it? (Z' score, how well can we separate positive controls from negative controls)

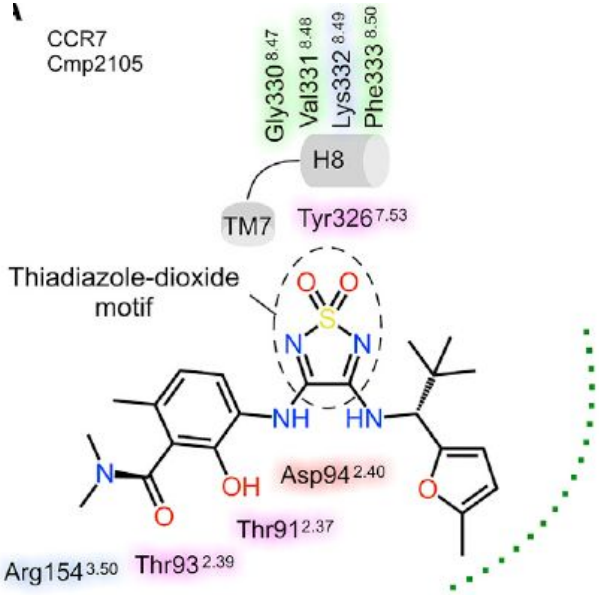
Q10: Why logarithm (usually base 10) transformation is often preferred to represent quantities such as IC_{50} and K_i? (presentation, as well as statistical mechanistics)

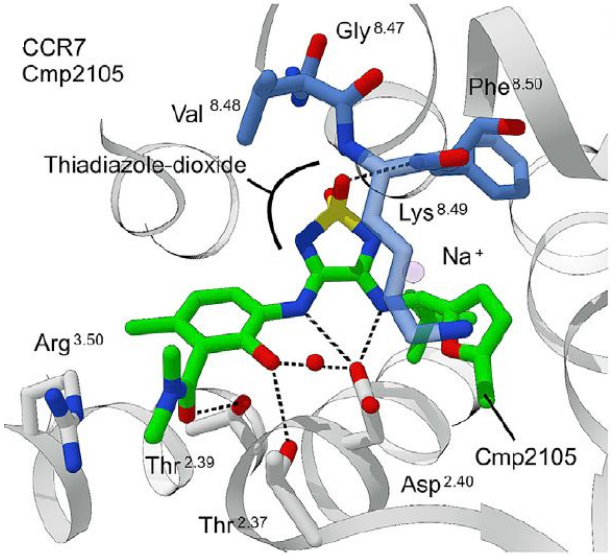
Questions from you:

1. The displacement method to measure ligand binding performance indirectly is not so clear to me.
2. I was a bit confused as to the statistical measure question.

AMIDD Lecture 6: Structure- and ligand-based drug design







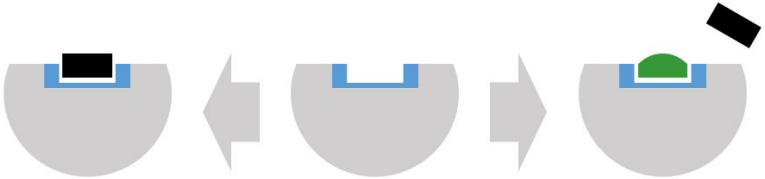
Dr. Jitao David Zhang, Computational Biologist

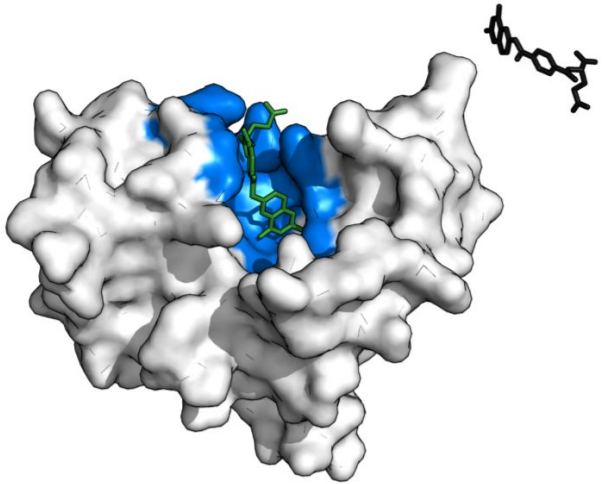
[1] Pharmaceutical Sciences, Pharma Research and Early Development, Roche Innovation Center Basel, F. Hoffmann-La Roche

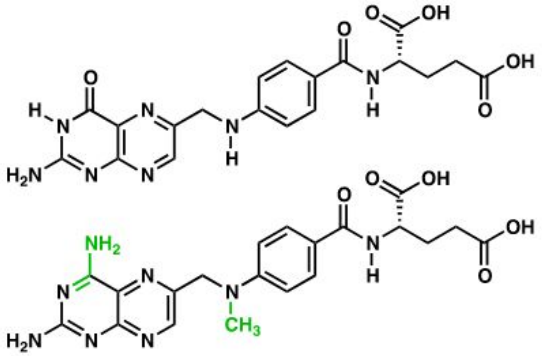
[2] Department of Mathematics and Informatics, University of Basel

Jaeger, Kathrin, Steffen Bruenle, Tobias Weinert, Wolfgang Guba, Jonas Muehle, Takuya Miyazaki, Martin Weber, et al. "Structural Basis for Allosteric Ligand Recognition in the Human CC Chemokine Receptor 7." Cell 178, no. 5 (August 22, 2019): 1222-1230.e10.

Seeing how a drug work







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.

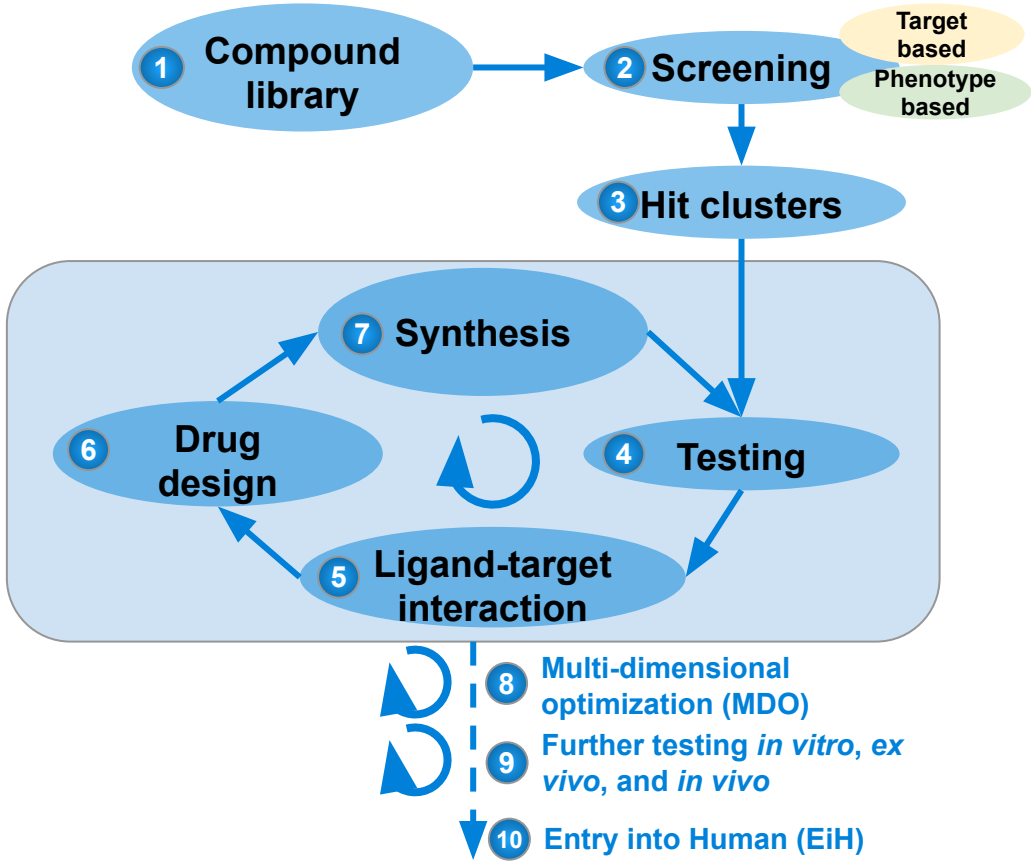
The natural substrate: dihydrofolic acid (vitamin B9), in black. Dihydrofolic acid is the *natural ligand* of DHFR.

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.

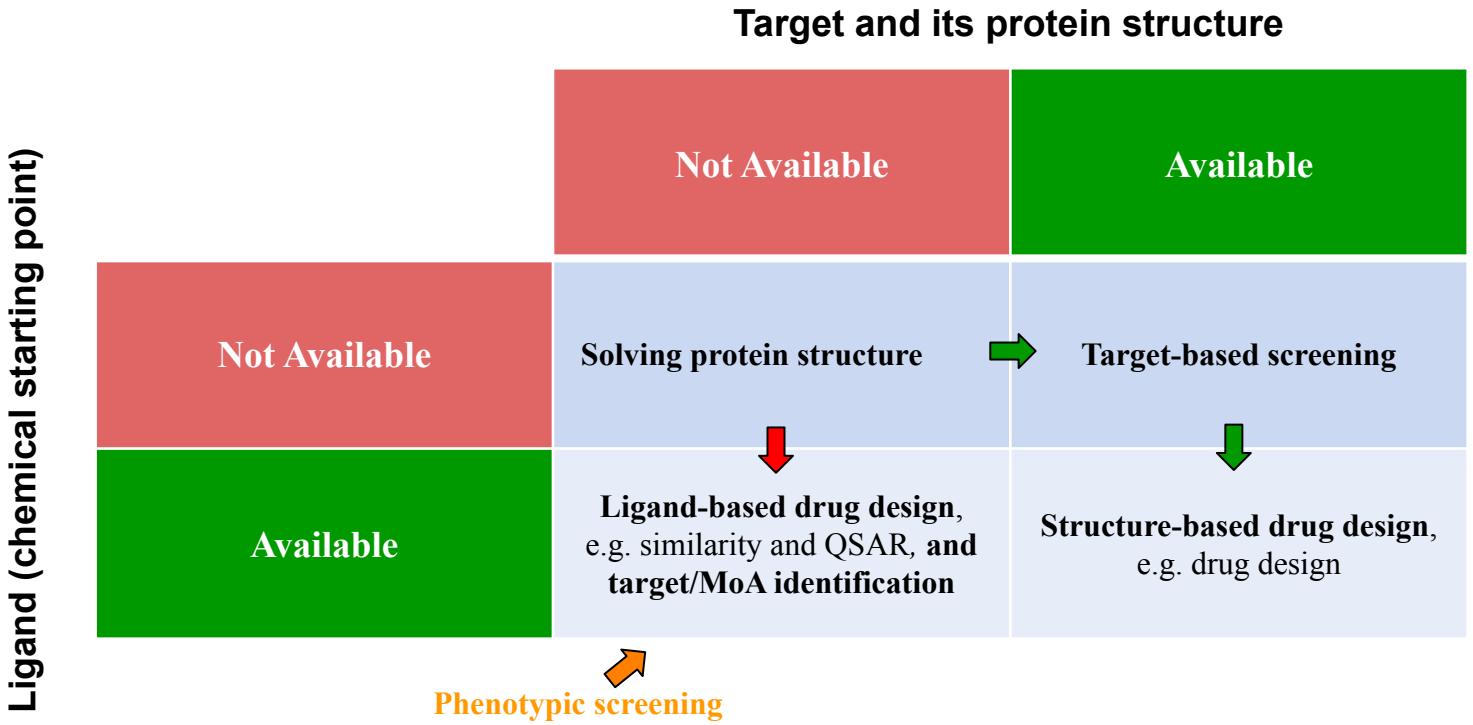
Workflow in a typical drug-discovery program

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).



A schematic presentation of structure-based drug discovery

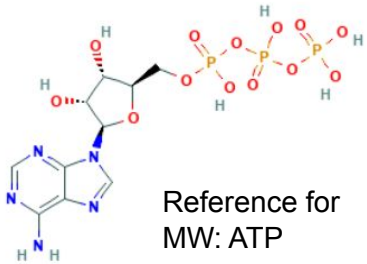
Structure-based and ligand-based drug design



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

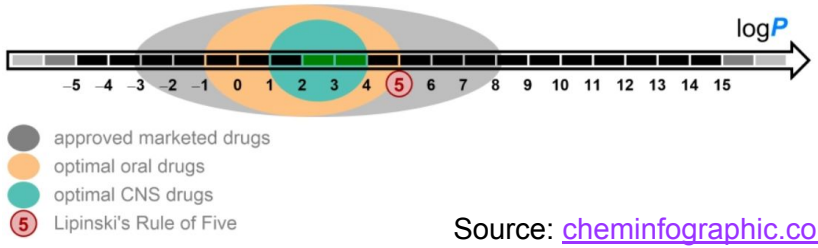
Lipinski's Rule of Five of small-molecule drugs as a rule of thumb

- **HBD<=5:** No more than **5 hydrogen-bond donors,** *e.g.* the total number of nitrogen–hydrogen and oxygen–hydrogen bonds.
- **HBA<=10**: No more than **10 hydrogen-bond acceptors**, *e.g.* all nitrogen or oxygen atoms.



Reference for MW: ATP

- **MW<500**: A **molecular weight** less than **500 Daltons, or** 500 g/mol. Reference: MW_{ATP}= ~507.
- **logP<=5**: An **octanol-water partition coefficient (log_{10} P)** that does not exceed **5.**



Source: cheminfographic.com

Many drugs make it by breaking RO5

Table 1. New FDA Approvals (2014 to Present)a 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

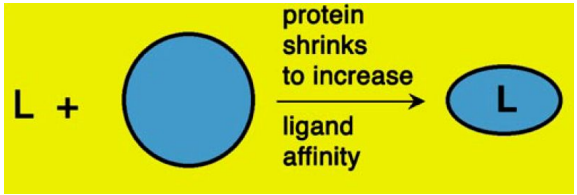
DeGoey, *et al.*. 2018. "Beyond the Rule of 5: Lessons Learned from AbbVie's Drugs and Compound Collection." Journal of Medicinal Chemistry 61 (7): 2636–51.

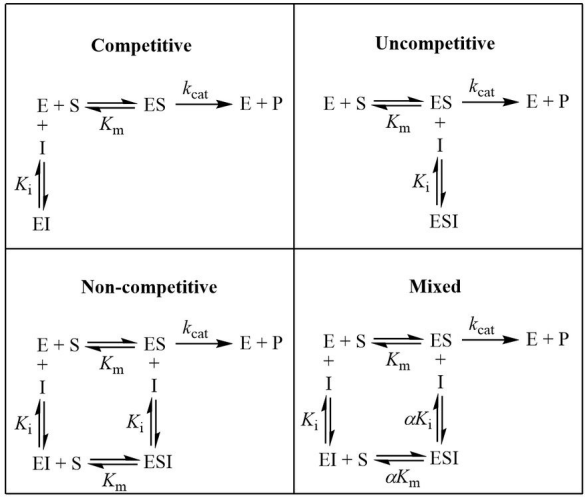
Outline

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

The 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 (ΔG=ΔH-TΔS).
  - 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}.





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

From the law of mass action to ligand-target interaction

$$\mathrm{L} + \mathrm{R} \underset{k_2}{\overset{k_1}{\rightleftharpoons}} \mathrm{LR}$$

The law of mass action

$$\frac{d[LR]}{dt} = k_1[L][R] - k_2[LR]$$

At equilibrium, no net change of [LR]

$$k_1[L][R] = k_2[LR]$$

$$R_{total} = [R] + [LR]$$

$$k_1[L]([R_{total}] - [LR]) = k_2[LR],$$

$$[LR] = \frac{k_1[L][R_{total}]}{k_1[L] + k_2}$$

$K_D \equiv k_2/k_1$

$$[LR] = \frac{[R_{total}][L]}{[L] + K_D}$$

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

Lock and key

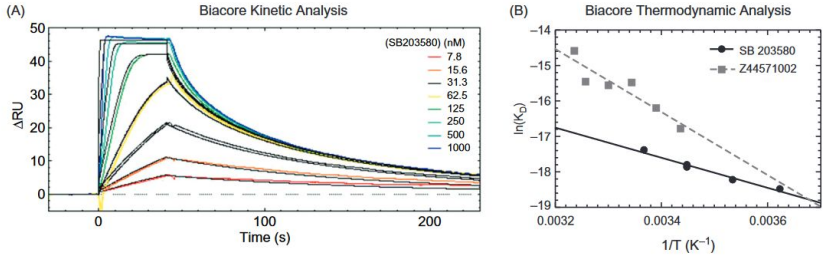
Gibbs free energy

$$\Delta G = \Delta H - T\Delta S$$

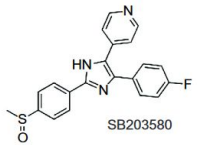
Van't Hoff Equation

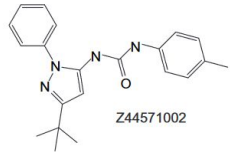
$$\Delta G = -RT \ln K$$

$$ln(K_D) = \frac{\Delta H}{R}\frac{1}{T} - \frac{\Delta S}{R}$$



Compound Name	k1	k2	KD	ΔG	ΔH	TΔS
Z44571002	2.2e4±3e2	0.001±8.0e6	5.2e−8	−40	−75	−35
SB203580	1.7e6±1.7e5	0.130±0.014	7.8e−8	−43	−36	7.5

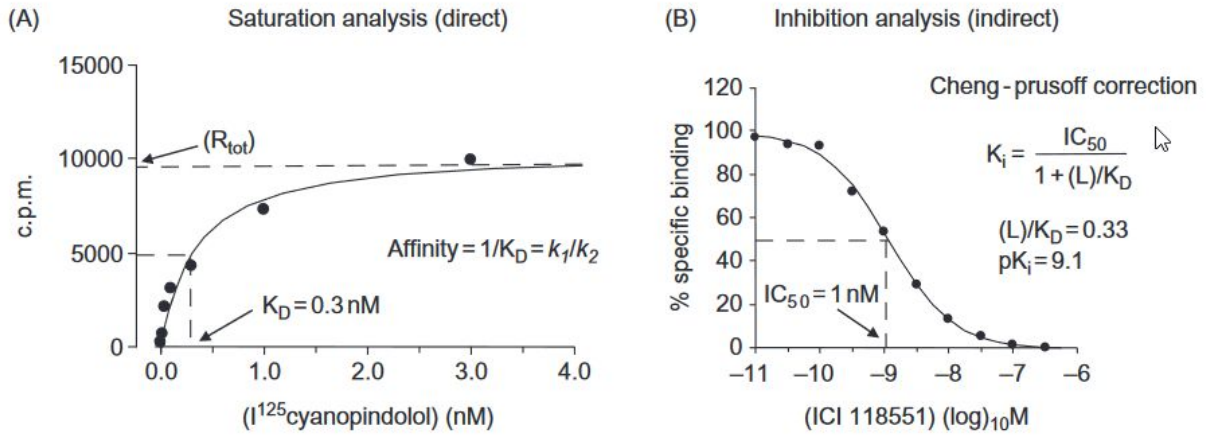




Kinetic and thermodynamic measurements of two p38α inhibitors. (A) The time course of SB203580 binding to immobilized mitogen activated kinase p38α. The y-axis shows the mass change resulting from compound binding to p38α. At t=0, a range of SB203580 concentrations were passed across the immobilized 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 (KD(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 TΔS 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 β2-adrenoceptor ligand. The curve describes a rectangular hyperbola which saturates at high ligand concentration. The ligand dissociation constant (KD) was estimated as 0.3 nM and is a measure of the ligand affinity. (B) A typical inhibition analysis using membranes expressing the human β2-adrenoceptor and employing 0.1 nM I^{125} cyanopindolol as the labeled ligand. The displacing ligand, the selective β2-adrenoceptor antagonist ICI 118551, produces complete inhibition of the specific binding yielding an IC50 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

$$\mathrm{E} + \mathrm{S} \underset{k_r}{\overset{k_f}{\rightleftharpoons}} \mathrm{ES} \xrightarrow{k_{cat}} \mathrm{E} + \mathrm{P}$$

The law of mass action

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

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

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

$K_D \equiv \frac{k_r}{k_f}$

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

$$v = \frac{d[P]}{dt} = k_{cat}[ES] = \frac{k_{cat}[E]_0[S]}{K_D + [S]}$$

$V_{max} \equiv k_{cat}[E]_0$

$$v = \frac{V_{max}[S]}{K_D + [S]}$$

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, …), the half-saturation constant.
- The Michaelis-Menten model is a special case of the Hill function ($n=1$).

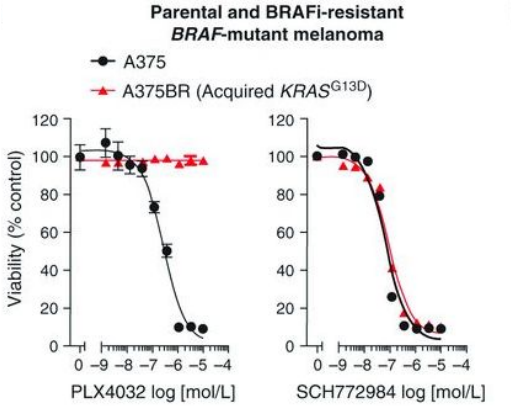
$$H = H_{max}\frac{x^n}{k^n + x^n}$$

The general form of the Hill function

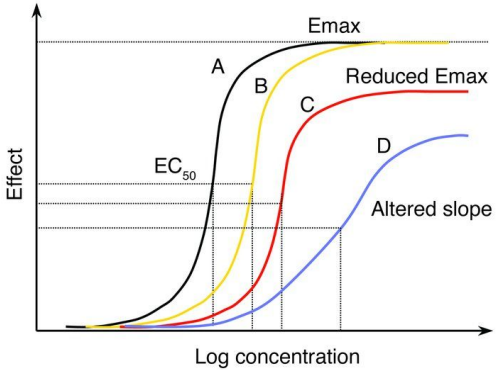
$$E = E_{max}\frac{[L]^n}{EC_{50}^n + [L]^n}$$

$$= E_{max}\frac{1}{1 + (\frac{EC_{50}}{[L]})^n}$$

Modelling the dose-dependent effect



Morris *et al.* Cancer Discov; 3(7); 742–50. ©2013 AACR.

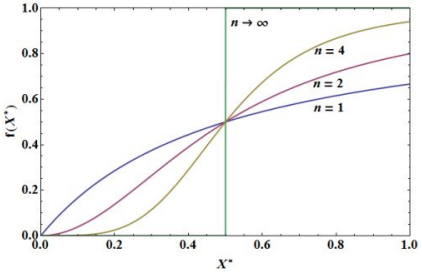


White. *J Clin Invest.* 2004;113(8):1084-1092. https://doi.org/10.1172/JCI21682.

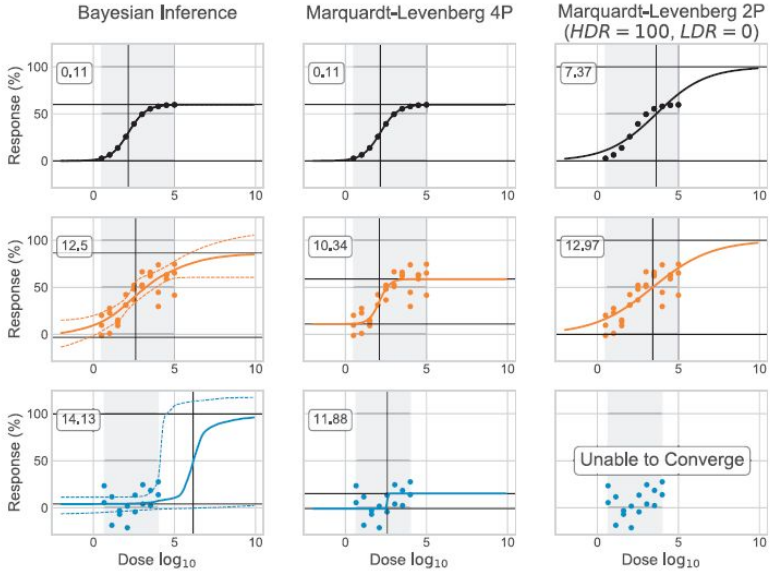
Suppose it is an antiviral drug, compared with curve B, what does curve A, C, and D suggest?

Theoretical and practical considerations about the Hill function

- 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.
- Dose-response data may look quite different from the ideal curve (bottom panel). By using a Bayesian inference approach (setting priors), it is possible to perform inference even with ill-looking data.



From the biophysics wiki article by Andreas Piehler



The Bayesian inference approach versus the non-Bayesian Marquardt-Levenberg algorithm for non-linear regression fitting (other options are gradient descent and Gauss-Newton methods). 4P: four parameter model; 2P: two parameter model (IC50 and *n*). 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.

Four classical classes of mathematical models

Compartment models

$$\frac{d[LR]}{dt} = k_1[L][R] - k_2[LR]$$

Kinetics of ligand-target interaction

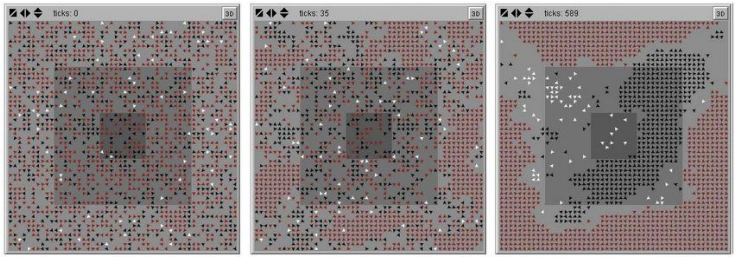
$$\frac{dx}{dt} = \alpha x - \beta xy,$$
$$\frac{dy}{dt} = -\gamma y + \delta xy,$$

The Lotka-Volterra equations modelling predator-prey relationships.

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

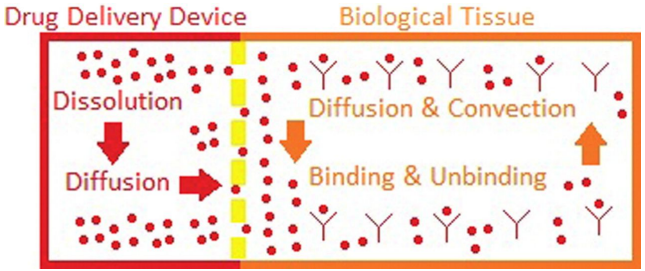
The SIR (S=susceptible, I=infectious, R=removed) model of epidemiology

Particle models



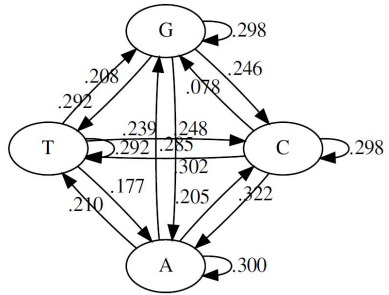
A Study on Socio-spatial Segregation Models Based on Multi-agent Systems by Quadros *et al*. (2012). 10.1109/BWSS.2012.14.

Transport models



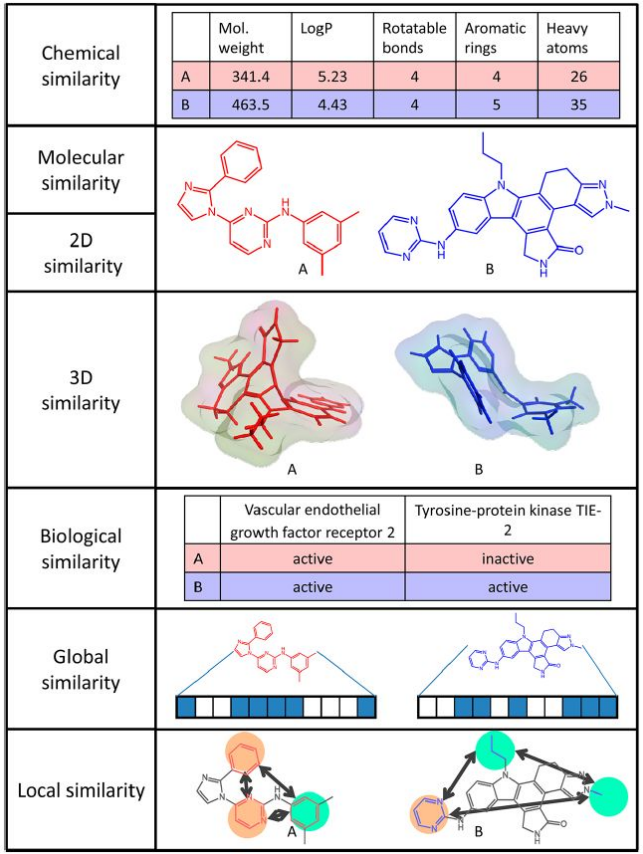
McGinty, Sean, and Giuseppe Pontrelli. 2015. "*A General Model of Coupled Drug Release and Tissue Absorption for Drug Delivery Devices*." Journal of Controlled Release 217 (November): 327–36.

Finite state models



A finite-state Markov chain modelling DNA sequences

Molecular similarity and similarity measures



	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

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

Table 2 Formulas for the various similarity and distance metrics

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

S denotes similarities, while *D* denotes distances. The two can be converted to each other by *similarity=1/(1+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.

Quantitative Structure-Activity Relationships (QSARs)

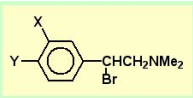
QSAR is a statistical modelling of correlation between biological activity and physicochemical properties, or $\Delta\phi=f(\Delta S)$, where ϕ indicates a biological activity and S indicates a chemical structure (1868-1869).

Molecular Descriptors (MD)

Compounds (C)	Target property	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~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.



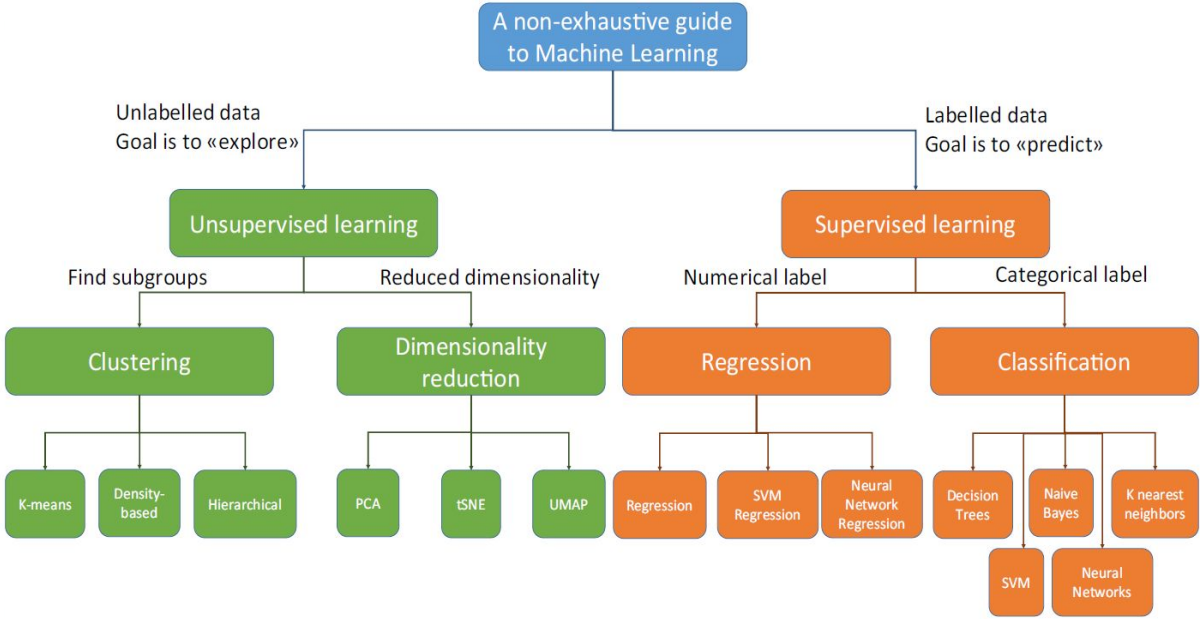
meta	para	meta-					para-					log 1/C	log 1/C
(X)	(Y)	F	Cl	Br	I	Me	F	Cl	Br	I	Me	obsd.	calc.a)
H	H											7.46	7.82
H	F						1					8.16	8.16
H	Cl							1				8.68	8.59
H	Br								1			8.89	8.84
H	I									1		9.25	9.25
H	Me										1	9.30	9.08
F	H	1										7.52	7.52
Cl	H		1									8.16	8.03
Br	H			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

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

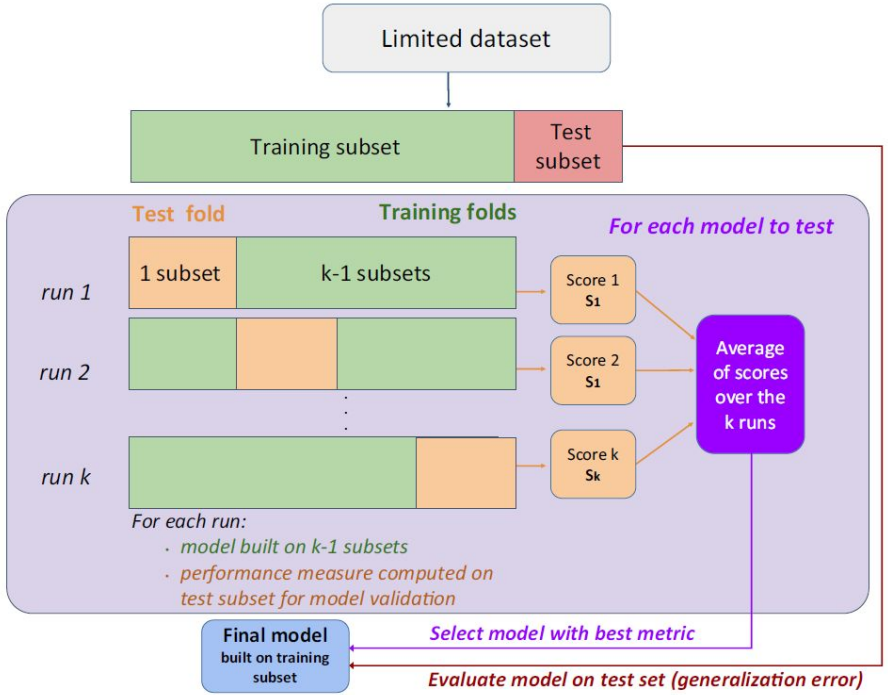
QSAR models mark the early adoption of statistical modelling and machine learning in drug discovery, the fifth type of mathematical modelling

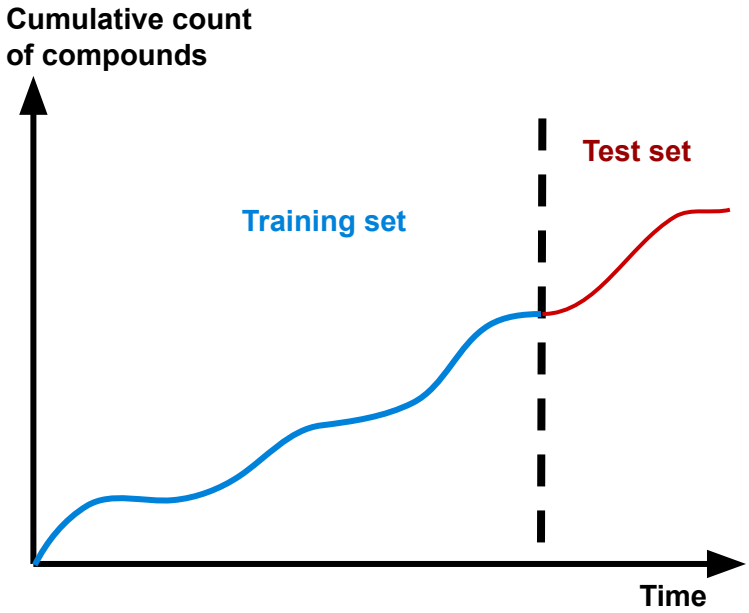
- QSAR is among the earliest subjects that used machine learning and pattern recognition in drug discovery.
- **Advantages**: technically easy, fast, and many models are useful as filters.
- **Disadvantages**: statistical models cannot capture mechanistic aspects of biochemical interactions, limited 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*.

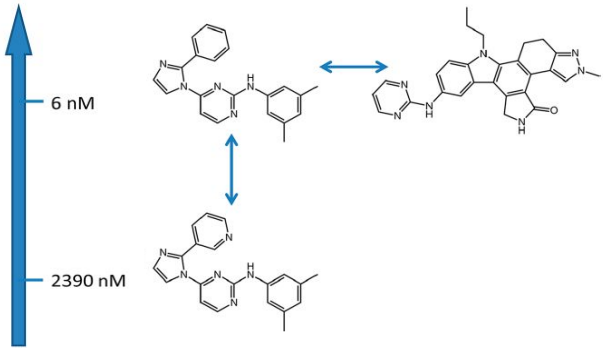
The general practice of training a supervised learning model

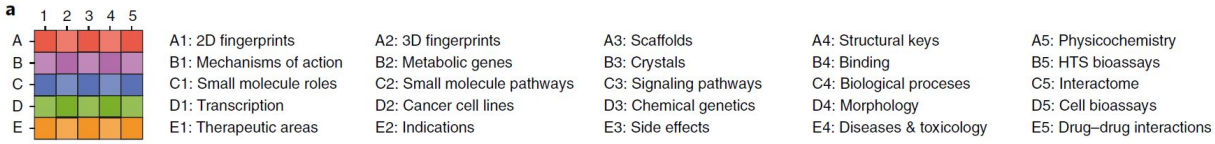




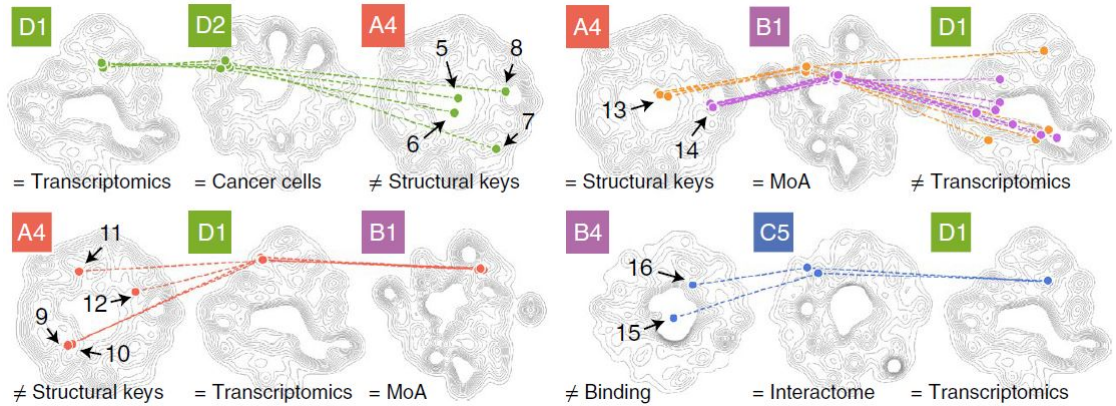
(Left) To assess the generalization ability of a supervised learning algorithm, data are separated into a training subset used for building the model and a test subset used to assess the generalization error (from Badillo *et al.*, 2020) (Right) Temporal validation is especially important for drug discovery,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

Molecular similarity does not equal biological similarity





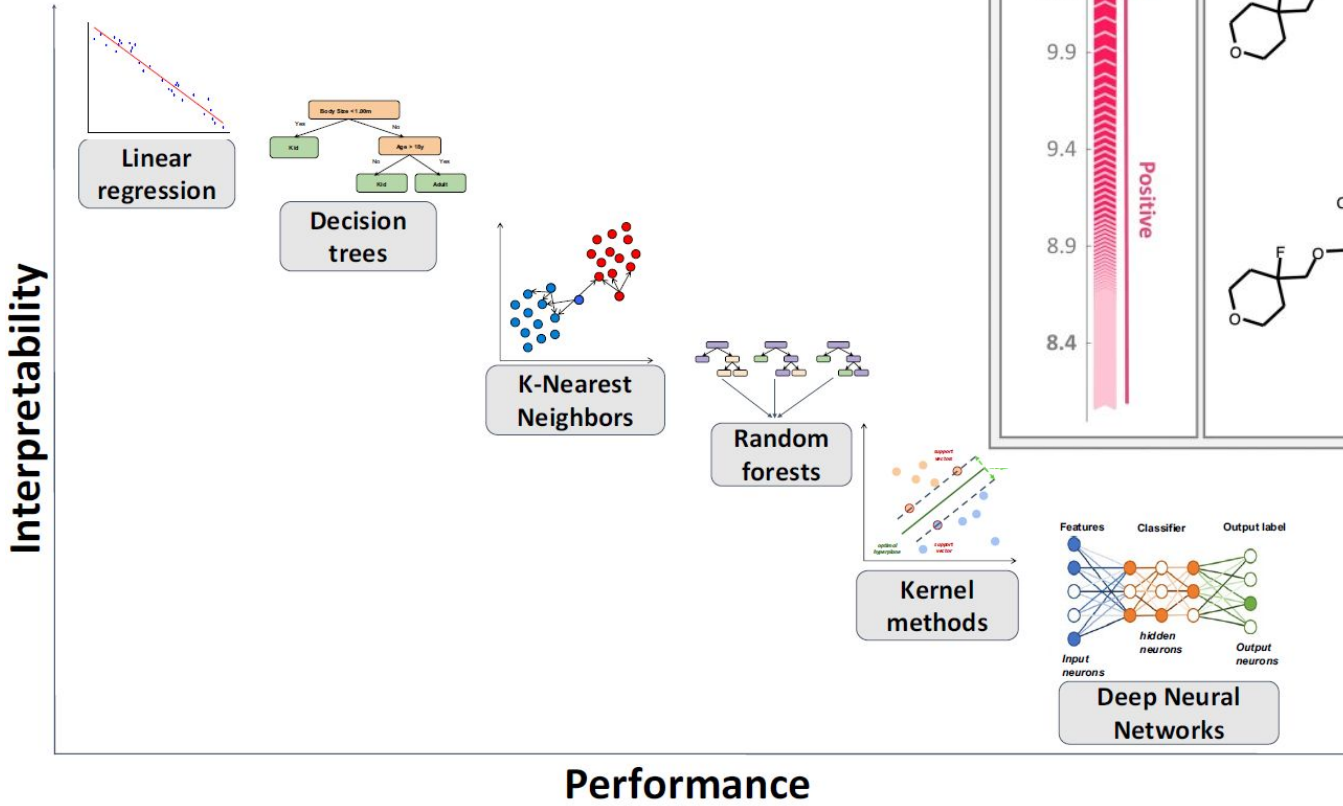
A: Chemistry
B: Targets
C: Biological network
D: Cells
E: Clinical readout

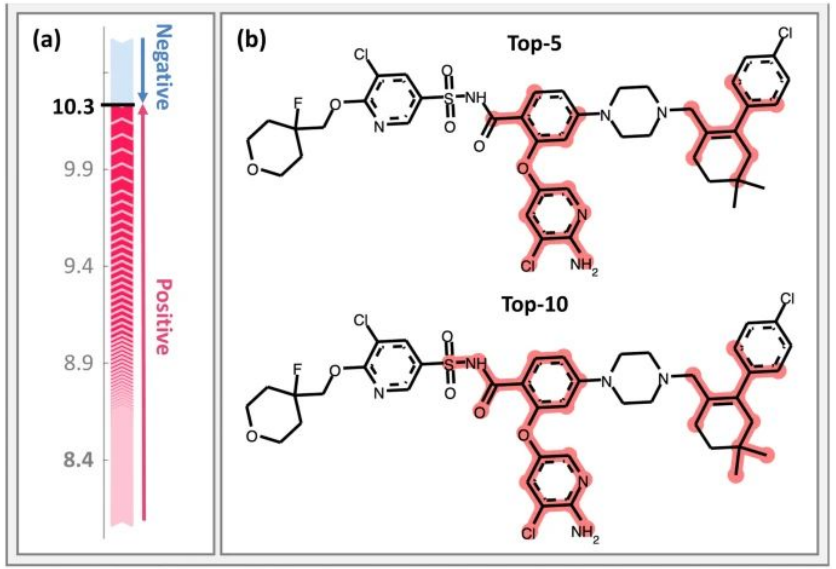


Watch out biological activity cliffs!
Similarity does not imply activity. Three vascular endothelial growth factor receptor 2 (VEGFR2) ligands are shown that represent different similarity−activity relationships.

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.

Interpretable and Causal Models will become more important

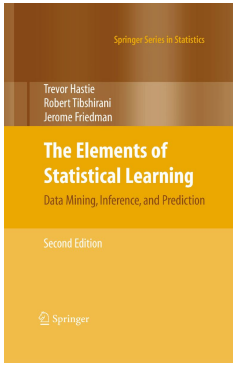


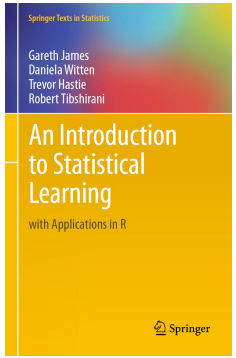


Rodríguez-Pérez, Raquel, and Jürgen Bajorath. "Interpretation of Machine Learning Models Using Shapley Values: Application to Compound Potency and Multi-Target Activity Predictions." Journal of Computer-Aided Molecular Design 34, no. 10 (October 1, 2020): 1013–26..

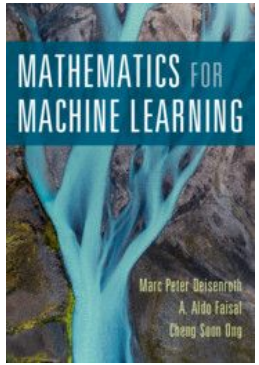
Resources for learning about machine learning

ESL and ISL: From a frequentist view (almost)

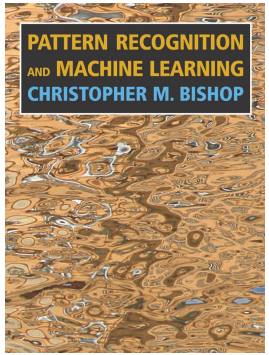


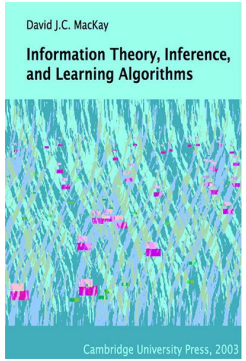


Mathematical foundations

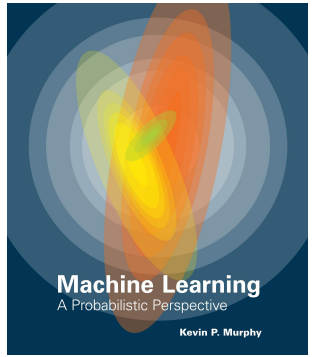


PRML and ITILA: From a Bayesian view

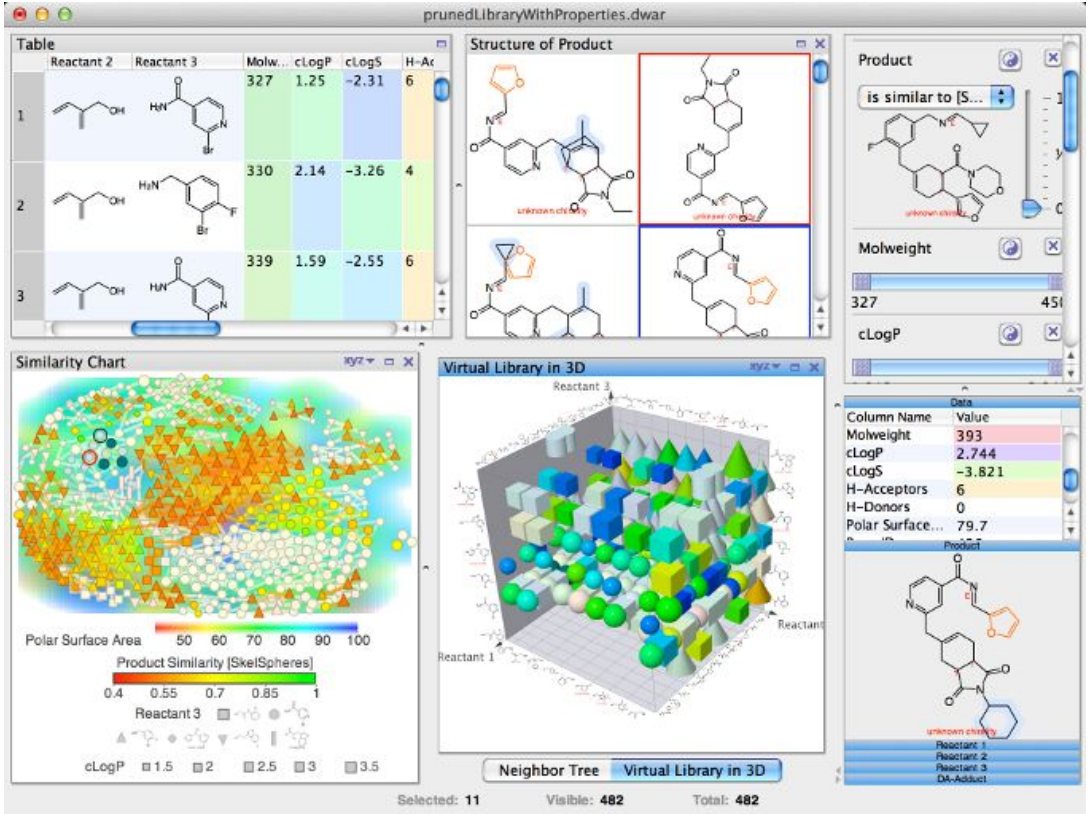




MLaPP: Application oriented, more accessible, and balanced views



DataWarrior: an open-source program for data visualization and analysis with chemical intelligence



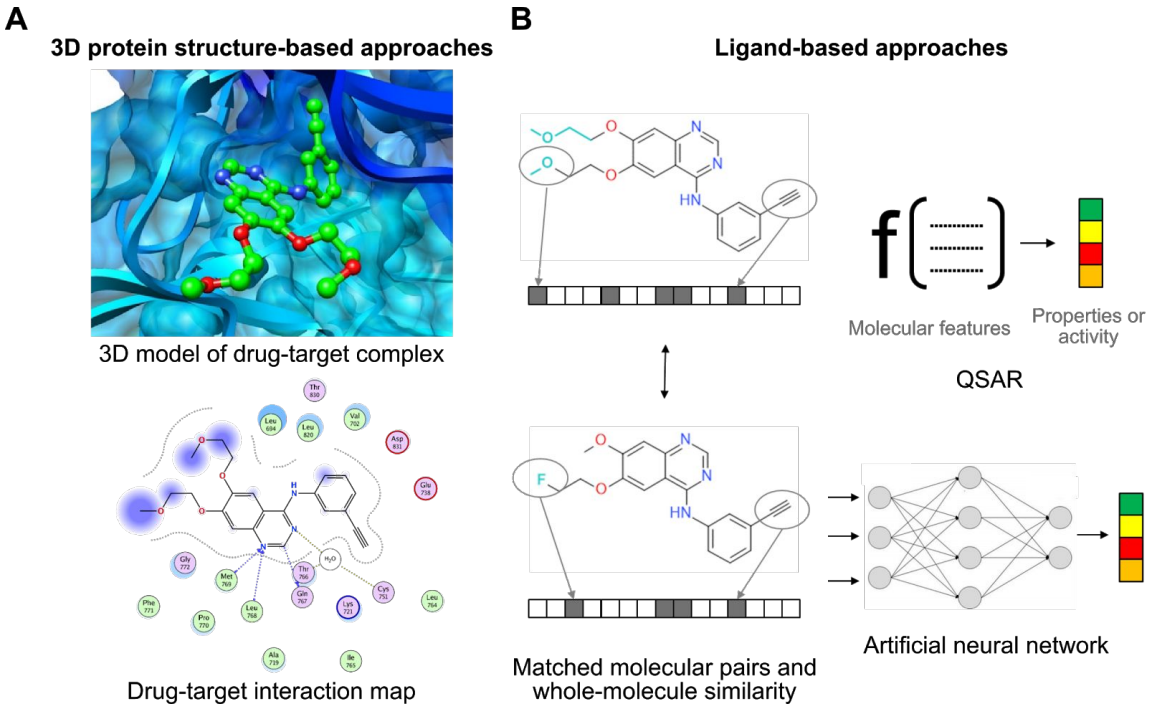
DataWarrior was and still is developed at Actelion/Idorsia Pharmaceuticals Ltd.

Selected subset of functionalities

- Molecular descriptor calculation
- Similarity calculation
- Compound clustering
- Docking
- …

Thomas Sander, Joel Freyss, Modest von Korff, Christian Rufener. DataWarrior: An Open-Source Program For Chemistry Aware Data Visualization And Analysis. J Chem Inf Model 2015, 55, 460-473, doi 10.1021/ci500588j

Summary and Q&A



Today we learned ligand-target interaction and **molecular modelling techniques:**

- (A) 3D protein structure-based approaches. An example with docking can be found in the backup slides.
- (B) Ligand-based approaches (similarity search). Another example of amphiphilicity can be found in the backup slides.

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.

Offline activities

- Read selected pages of *Computational Methods in Drug Discovery* by Sliwoski *et al*. Please submit your results to the Google Form.
- Fill the anonymous survey #6.
- Optional reading: Badillo *et al.* 2020. "An Introduction to Machine Learning." Clinical Pharmacology & Therapeutics.

Resources and backup slides

Summary of basic concepts

- **Ligand:** the binding partner of a macromolecule (often proteins), for instance other proteins (in case of protein-protein inaction), substrates and allosteric modulators (in case of enzymes). Many drugs are ligands of proteins.
- **Binding:** the formation of interactions between a protein and its ligand. In drug discovery, we encounter more often transient and non-covalent interaction (i.e. no sharing of electrons between atoms), but there are drugs form reversible or irreversible covalent bonds.
- **Non-covalent interaction**: electromagnetic interactions between molecules or within a molecule without forming a chemical bond, *i.e.* no sharing of electrons between atoms. Non-covalent interactions are classified into four categories: *electrostatic*, *van der Waals forces*, *hydrophobic effects*, and *π-effects*. See Wikipedia for more details of these interactions.
- **Conformational change**: ligand binding often triggers a change in the shape of the protein, which alters its cellular function
- **Agonist *versus* antagonist**: an *agonist* activates the function of its target by binding, and an *antagonist* blocks the action of the target by binding.
- **Active site versus allosteric site:** active site is where the enzyme-substrate interaction happens, example: at the active site oxygen binds to heme, and CO can compete with oxygen for heme binding. Allosteric site (i.e. regulatory site) is any other site than the active site where a ligand can bind to modulate the protein function.

More about molecular interactions and drug design

- Molecular interactions for drug discovery
  - Bissantz, Caterina, Bernd Kuhn, and Martin Stahl. “A Medicinal Chemist’s Guide to Molecular Interactions.” Journal of Medicinal Chemistry 53, no. 14 (July 22, 2010): 5061–84. https://doi.org/10.1021/jm100112j. A comprehensive introduction to common types of interactions, their applications, and caveats of blindly following rules in drug design.
  - Persch, Elke, Oliver Dumele, and François Diederich. “Molecular Recognition in Chemical and Biological Systems.” Angewandte Chemie International Edition 54, no. 11 (2015): 3290–3327. https://doi.org/10.1002/anie.201408487. A comprehensive introduction to molecular recognition.
- How drug design help with drug discovery: ten real-life stories
  - Kuhn, Bernd, Wolfgang Guba, Jérôme Hert, David Banner, Caterina Bissantz, Simona Ceccarelli, Wolfgang Haap, et al. “A Real-World Perspective on Molecular Design.” Journal of Medicinal Chemistry 59, no. 9 (May 12, 2016): 4087–4102. https://doi.org/10.1021/acs.jmedchem.5b01875. The common themes summarized in the Conclusion are helpful in my opinion for any scientist working in quantitative aspects of drug discovery: (1) value of qualitative statements, (2) shaping chemical space, (3) the principle of parsimony, (4) annotation is half the battle, and (5) staying close to experiment.

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.

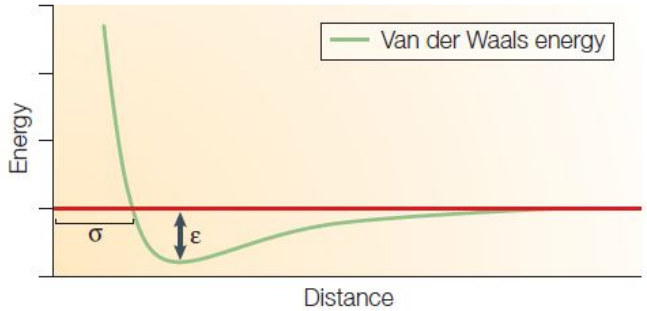
$$E_{coul}(r) = \sum_{i=1}^{N_A} \sum_{j=1}^{N_B} \frac{q_i q_j}{4\pi\varepsilon_0 r_{ij}}$$

Coulombic interactions (electrostatic interactions between electric charges)

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

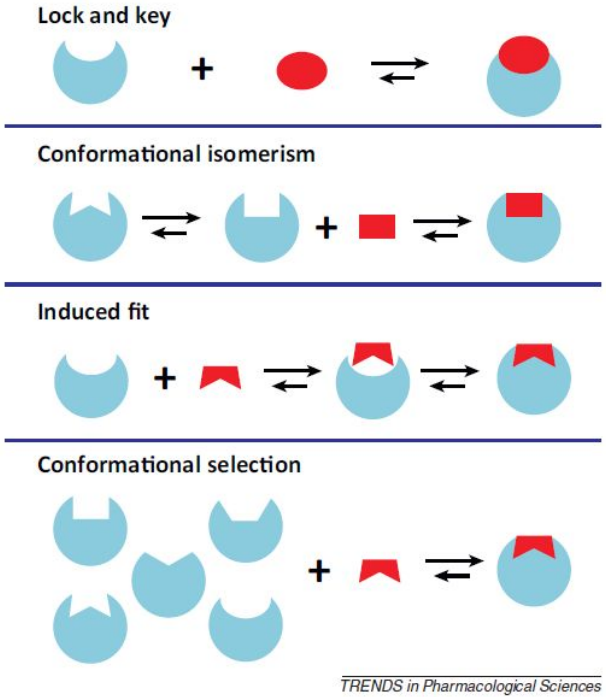
Lennard–Jones 12–6 function (intermolecular interactions without charge)

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

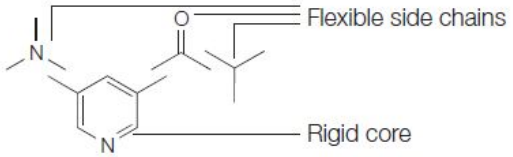


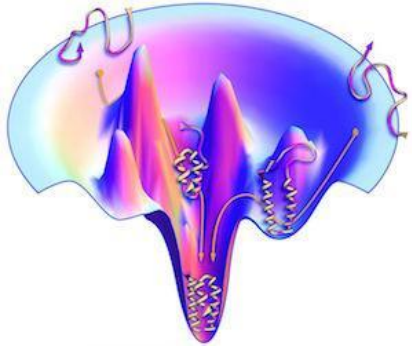
Kitchen, Douglas B., Hélène Decornez, John R. Furr, und Jürgen Bajorath. „Docking and Scoring in Virtual Screening for Drug Discovery: Methods and Applications". *Nature Reviews Drug Discovery* 3, Nr. 11 (November 2004): 935–49. https://doi.org/10.1038/nrd1549.

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.

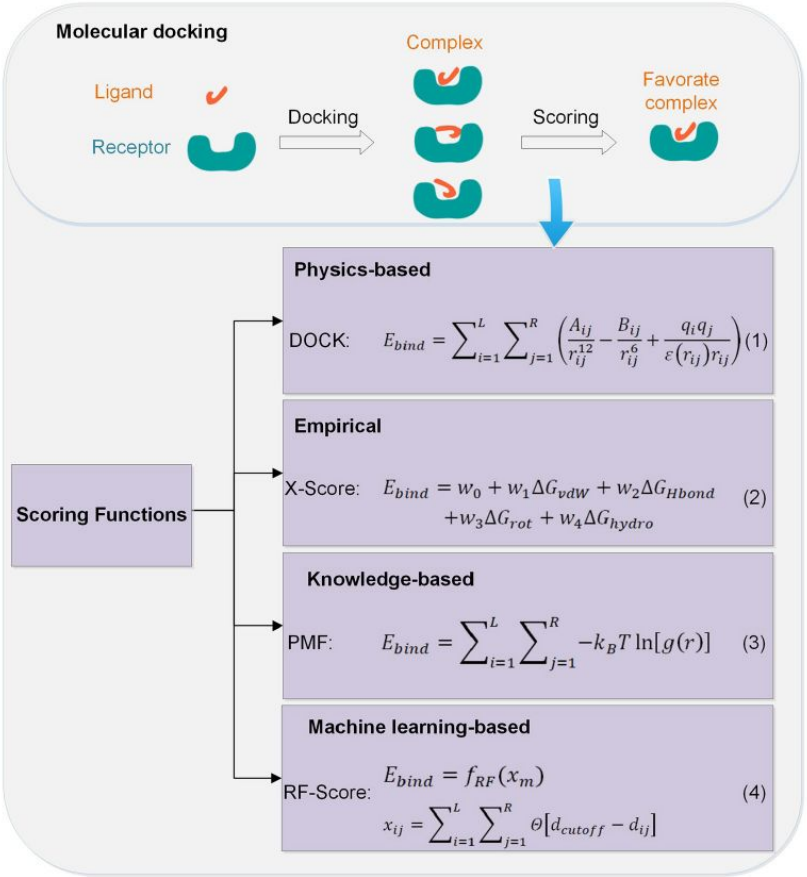


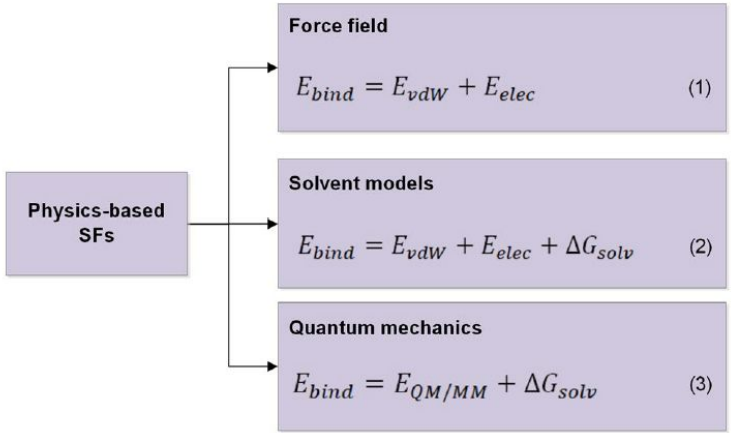


Methods to deal with ligand and protein flexibility

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

Types of scoring functions

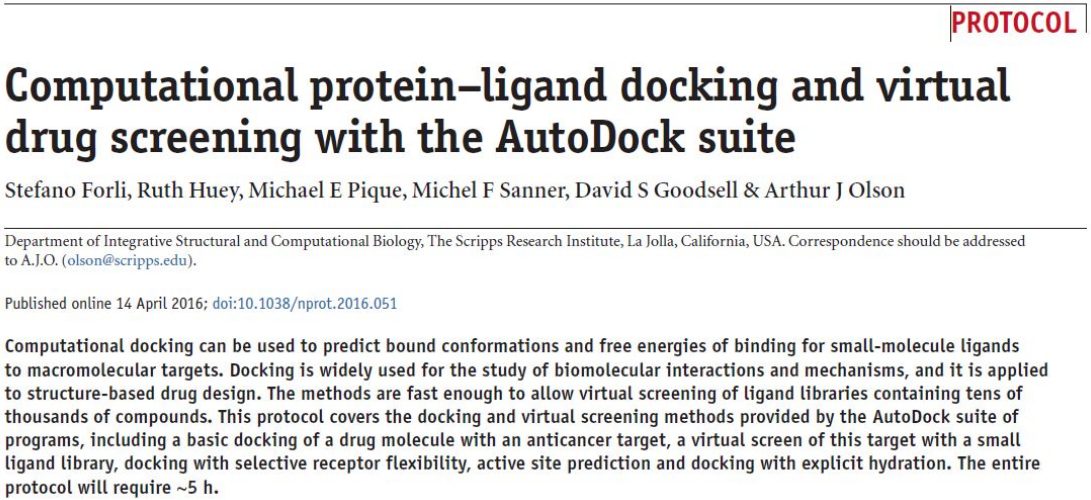




- Empirical scoring functions estimate the binding affinity of a complex by **summing up the important energetic factors for protein–ligand binding**, such as hydrogen bonds, hydrophobic effects, steric clashes, etc. It relies on training set and regression analysis.
- Knowledge-based scoring functions 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

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.

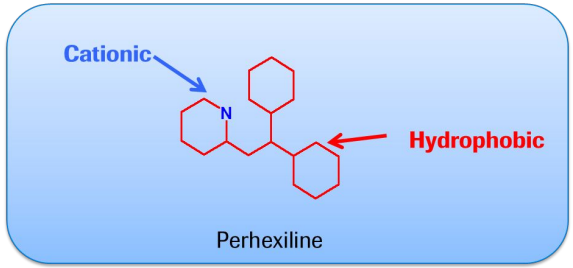
- Try docking yourself by following this protocol: Forli, Stefano, Ruth Huey, Michael E. Pique, Michel F. Sanner, David S. Goodsell, und Arthur J. Olson. „Computational Protein–Ligand Docking and Virtual Drug Screening with the AutoDock Suite“. *Nature Protocols* 11, Nr. 5 (Mai 2016): 905–19. https://doi.org/10.1038/nprot.2016.051.
- 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.**

More about the the Free-Wilson analysis

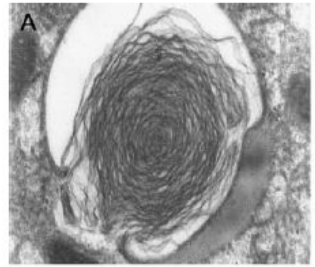
- *A Mathematical Contribution to Structure-Activity Studies* by Spencer M. Free and James W. Wilson, Journal of Medicinal Chemistry, 1964, and reviewed by Kubinyi, 1988.
- A Python implementation on GitHub, and a blog post going through examples, is shared by Pat Walters.
- Free-Wilson nonadditivity 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)

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.



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.

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.

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

$\vec{A}$: Caculated amphiphilic moment

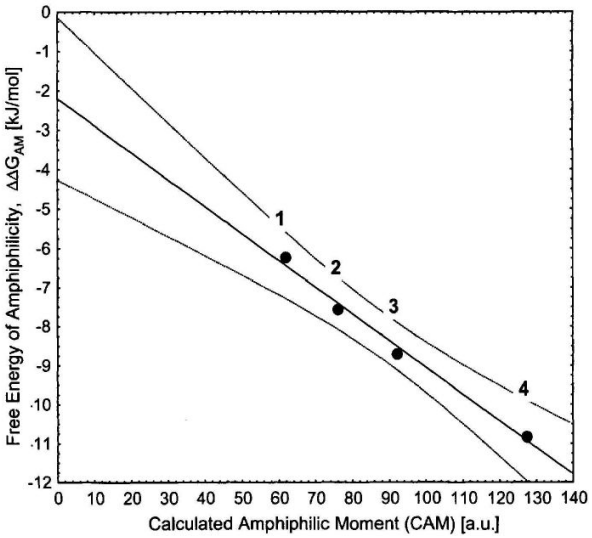
d: distance between the center of gravity of the charged part of a molecule and the hydrophobic/hydrophilic remnant of the molecule

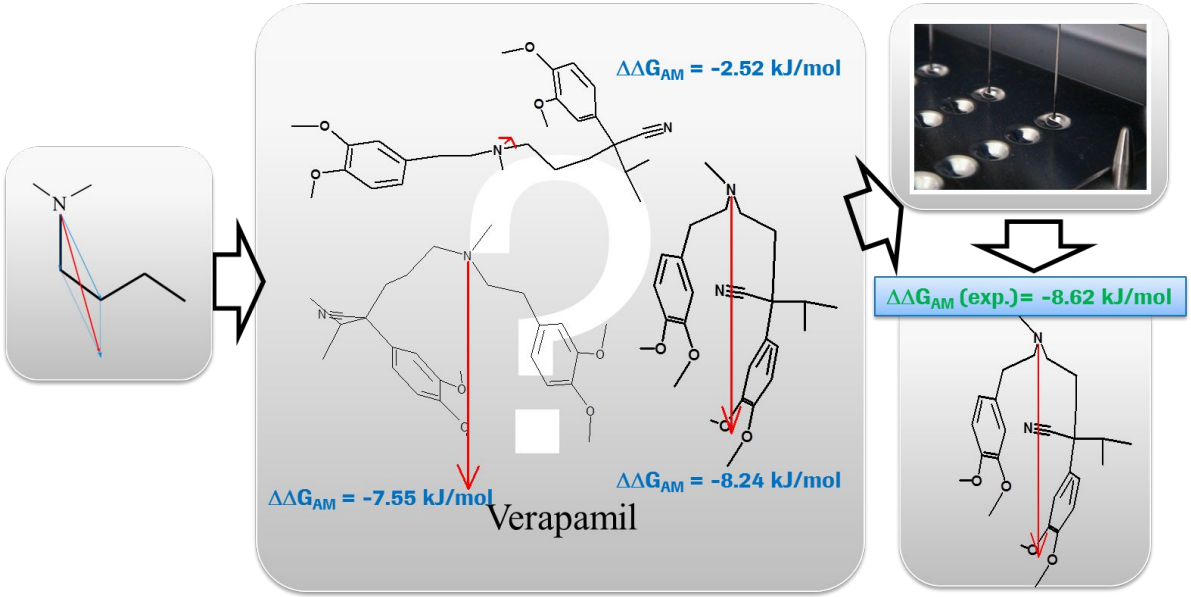
$\vec{\alpha}_i$: the hydrophobic/hydrophilic contribution of atom/fragment *i*

***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)

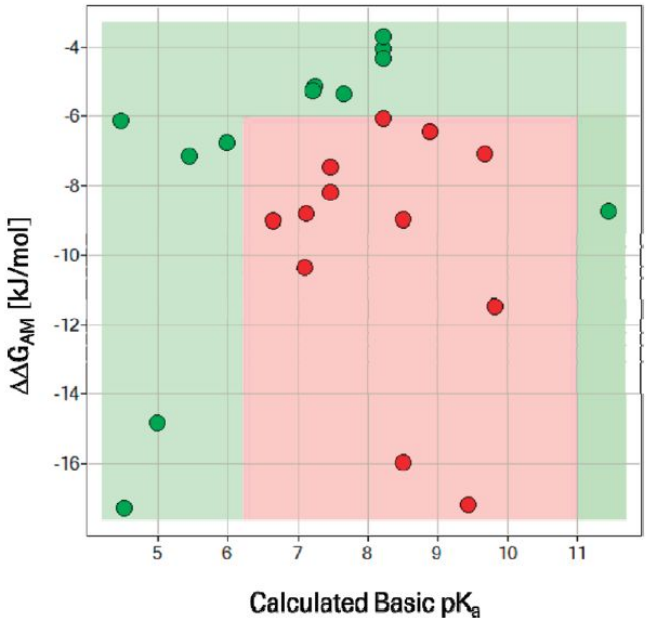




Iterative model building, experimentation, and model refining led to the predictive tool CAFCA

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 (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, natrium 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!**

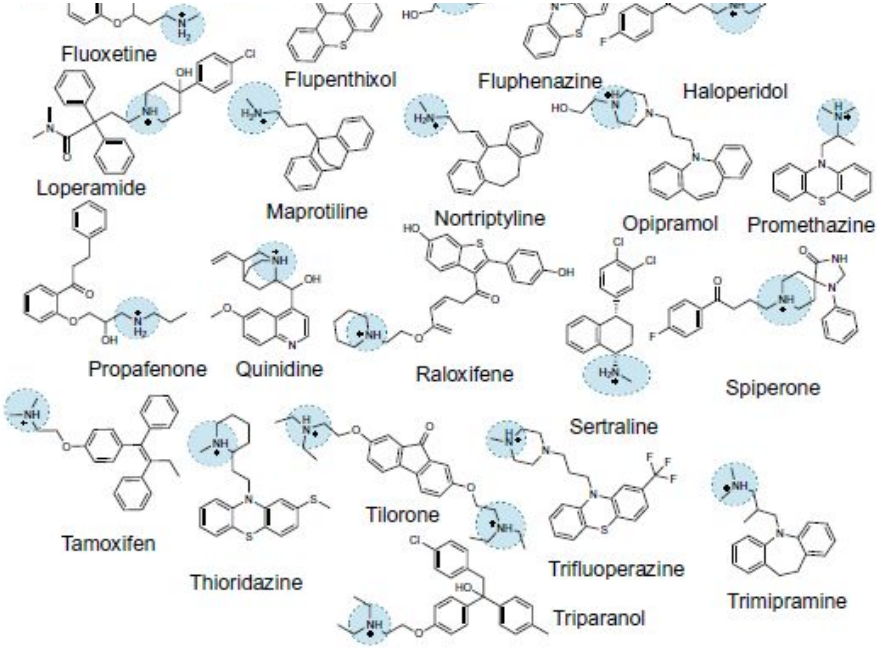
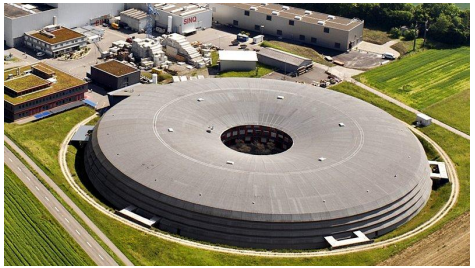


Fig. 1. Representative examples of CADs that are identified in SARS-CoV-2 drug repurposing screens.

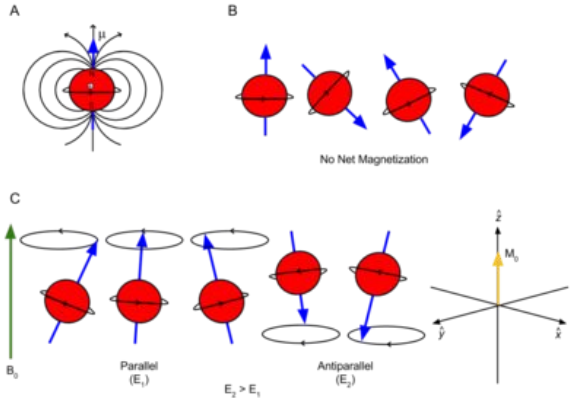
Tummino, Tia A., Veronica V. Rezelj, Benoit Fischer, Audrey Fischer, Matthew J. O'Meara, Blandine Monel, Thomas Vallet, et al. "Drug-Induced Phospholipidosis Confounds Drug Repurposing for SARS-CoV-2." Science 373, no. 6554 (July 30, 2021): 541–47. https://doi.org/10.1126/science.abi4708.

Resources about the mathematics underlying molecular structure determination

- **Mathematical and physical foundations**
  - **Recommended reading**: Mathematical techniques used in biophysics
  - Background on imaging physics at xrayphysics.com
  - Physics for life-science students at U Maryland
- **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*

NEXUS/Physics
Physics for life-science students