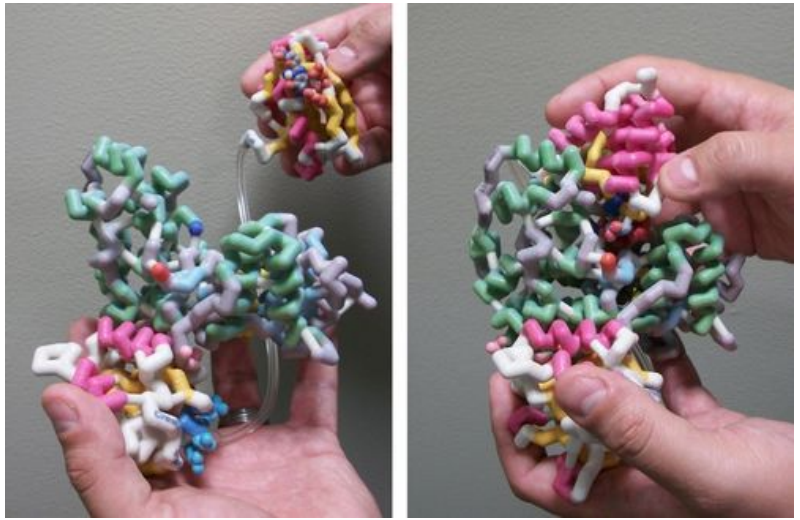
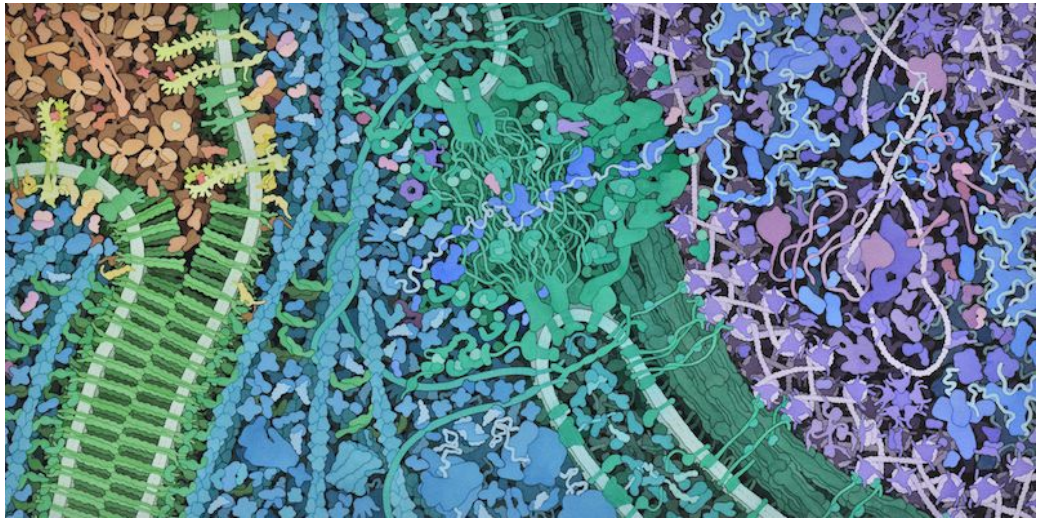
AMIDD Lecture 9: From individual interactions to networks





Left: A physical model of DUSP5 (bottom) binds to ERK2 (top), which is one interaction in the biological network induced by VEGF signaling. Right: 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

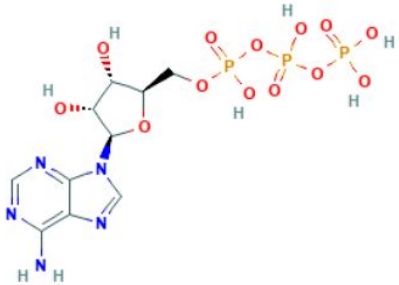
[2] Department of Mathematics and Informatics, University of Basel

Topics

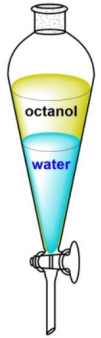
- Ligand-based and structure-based drug discovery
- Thermodynamic and kinetic views of ligand-protein binding
- From individual interactions to biological networks

Selected commonly used molecular descriptors

Molecular Weight (MW). for example, adenosine triphosphate (ATP), the *energy molecule*, has a MW of 507.

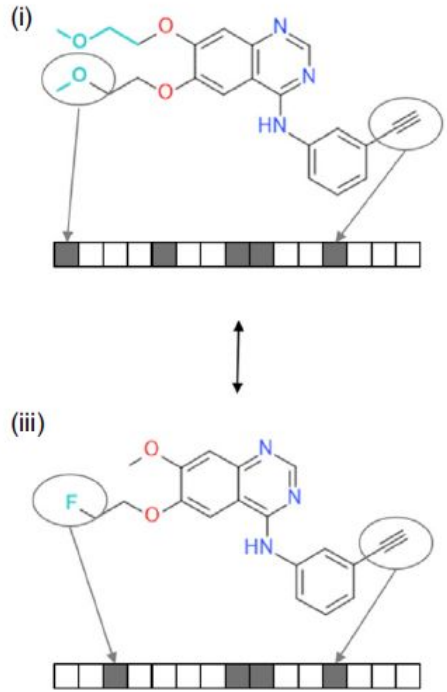


logP (partition coefficient) quantifies the hydrophilicity and hydrophobicity of a molecule. The calculated version (cLogP) exists as well.

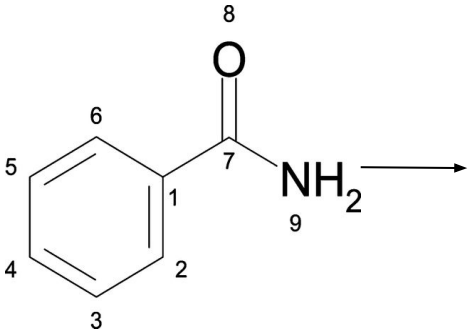


$$\log P_{\mathrm{oct/wat}} = \log\left(\frac{[\mathrm{solute}]^{\text{un-ionized}}_{\text{octanol}}}{[\mathrm{solute}]^{\text{un-ionized}}_{\text{water}}}\right)$$

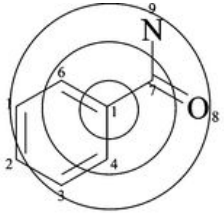
Molecular fingerprints: a set of techniques to represent molecules in a bit array.



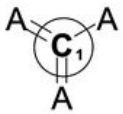
Extended-connectivity fingerprints (ECFPs) and Functional-class fingerprints (FCFPs) extract and compare (multi-)sets of subgraphs



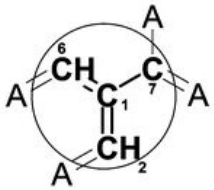
Benzamide



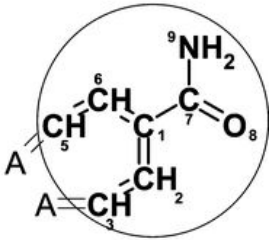
Considering atom 1 in benzoic acid amide



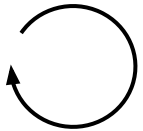
Iteration 0



Iteration 1



Iteration 2



hash(Daylight atomic invariants*)

*# non-H neighbors, bond order, atom number, atom mass, atom charge, attached Hs.

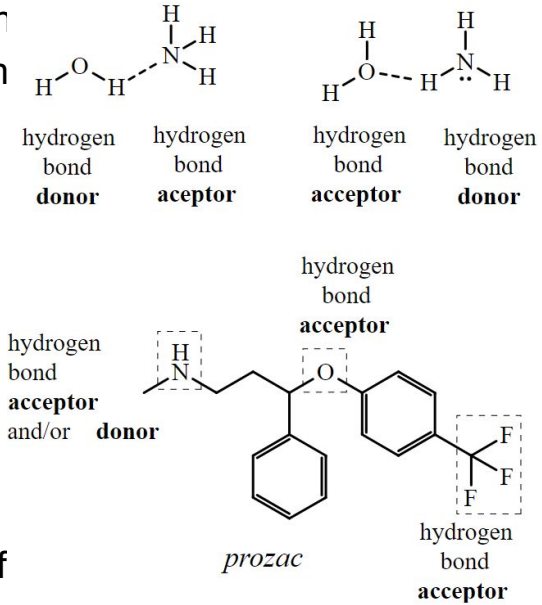
A bit array as output: *[0,0,0,0,0, …, 1, 0, 0,..., 1,0,...]*

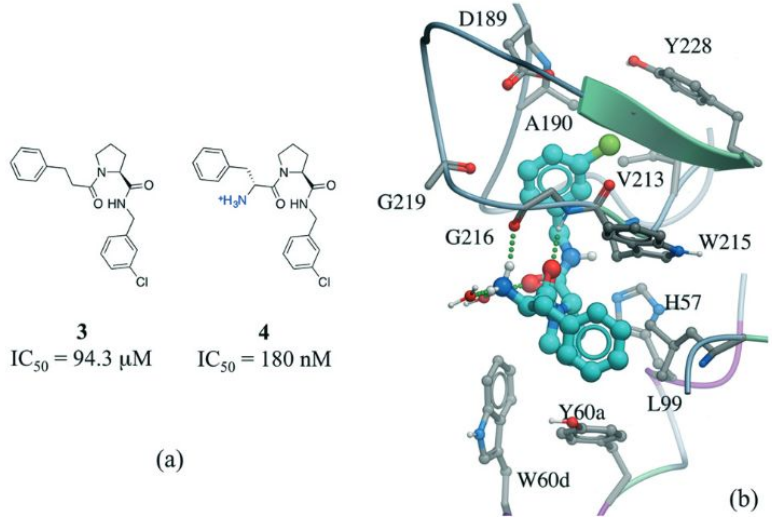
Implemented in RDKit and other software. Publication and tutorials: (1) Rogers, David, and Mathew Hahn. "Extended-Connectivity Fingerprints." Journal of Chemical Information and Modeling (2010). (2) Tutorial by Manish Kumar and (3) Tutorial by Leo Klarner.

Number of hydrogen bond acceptors and donors are important descriptors, too

A **hydrogen bond**: an electrostatic force of attraction between a hydrogen (H) atom which is covalently bonded to a more electronegative "donor" atom or group (Dn), and another electronegative atom bearing a lone pair of electrons—the hydrogen-bond acceptor.

Hydrogen bonds (H-bonds) both influence the structure of the molecule and its binding to the target.

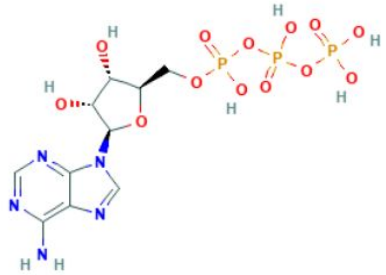




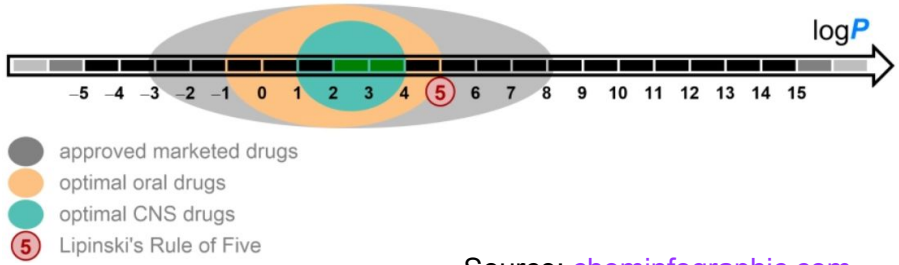
Effect of adding a hydrogen bond in a thrombin inhibitor: a) chemical structure of a pair of thrombin inhibitors; b) crystal structure of molecule 4 (cyan carbons) in complex with thrombin (PDB: 2ZC9). Hydrogen bonds are displayed in dotted green lines.

Lipinski's Rule of Five of small-molecule drugs

- **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
- **MW<500**: A **molecular weight** less than **500 Daltons, or** 500 g/mol.
- **logP<=5**: An **octanol-water partition coefficient (log P)** that does not exceed **5.** (10-based)



ATP (MW=507)



Source: cheminfographic.com

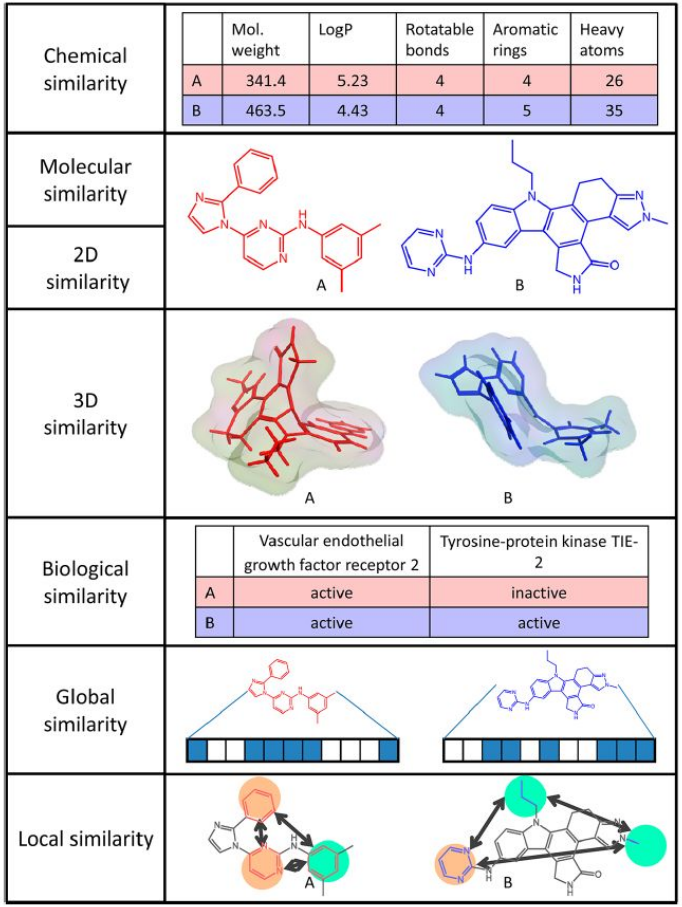
Rules are made to be broken: more drugs are now beyond the space of Ro5

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.

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

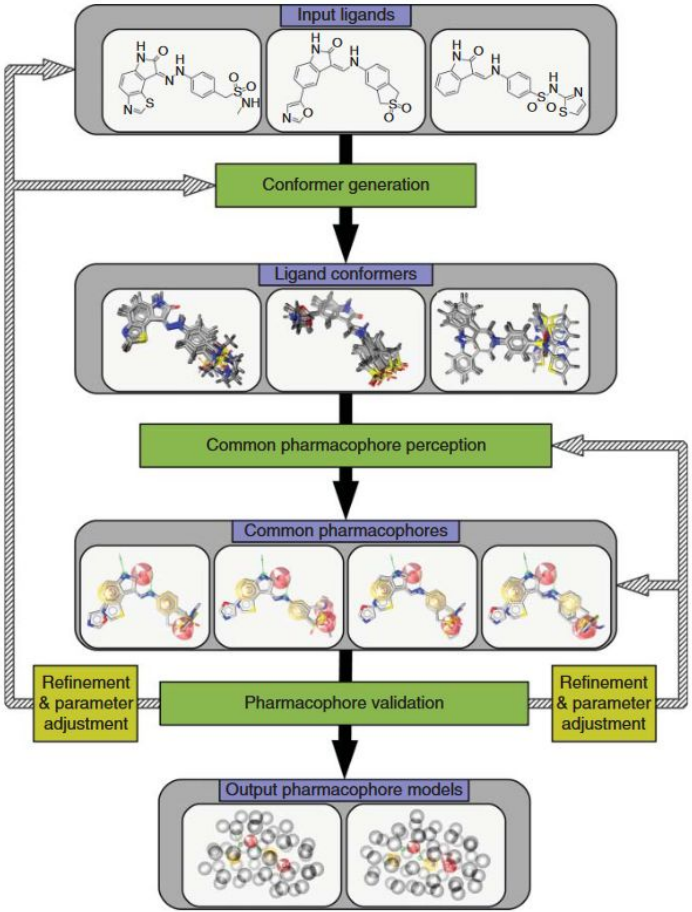
Table 2 Formulas for the various similarity and distance metrics

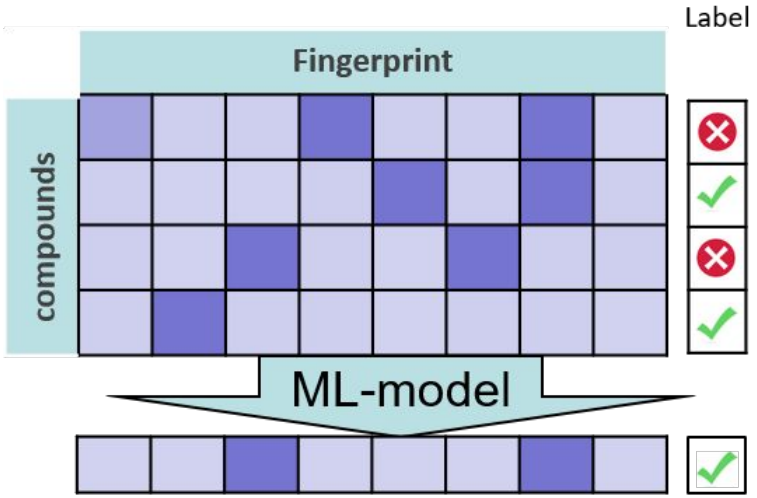
Distance metric	Formula for continuous variables[a]	Formula for dichotomous variables[a]
Manhattan distance	$D_{A,B} = \sum_{j=1}^{n} \lvert x_{jA} - x_{jB} \rvert$	$D_{A,B} = a + b - 2c$
Euclidean distance	$D_{A,B} = \left[\sum_{j=1}^{n} (x_{jA} - x_{jB})^2 \right]^{1/2}$	$D_{A,B} = [a + b - 2c]^{1/2}$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

S denotes similarities, while *D* denotes distances. The two can be converted to each other by *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.

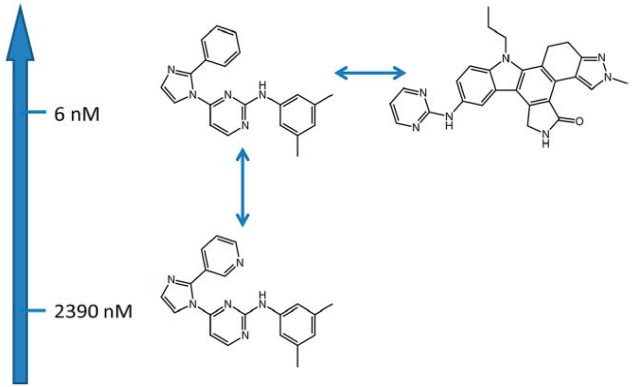
Pharmacore models and machine learning for ligand-based drug discovery





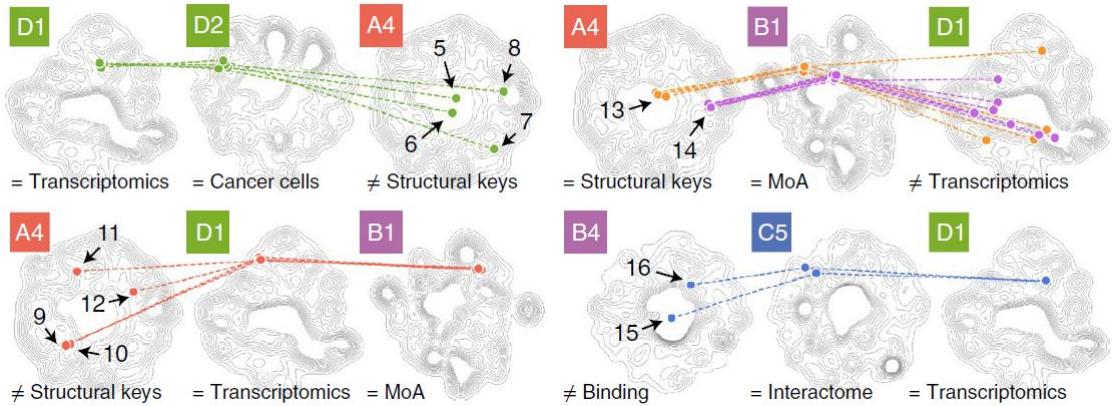
Left: Ligand-based pharmacophore modeling workflow, starting from a set of known active compounds. Right: Machine-learning model to predict the activity of unseen compound, starting from a set of known active and inactive compounds. Applied chemoinformatics: achievements and future opportunities. (Wiley-VCH, 2018), p271. TeachOpenCADD, T007, Ligand-based screening with machine learning

Molecular similarity does not equal biological similarity



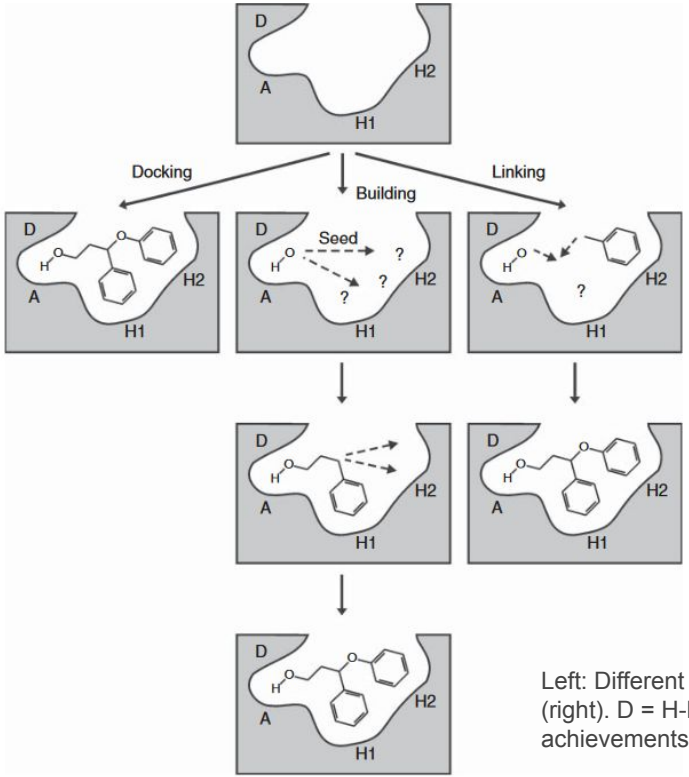
A: Chemistry
B: Targets
C: Biological network
D: Cells
E: Clinical readout

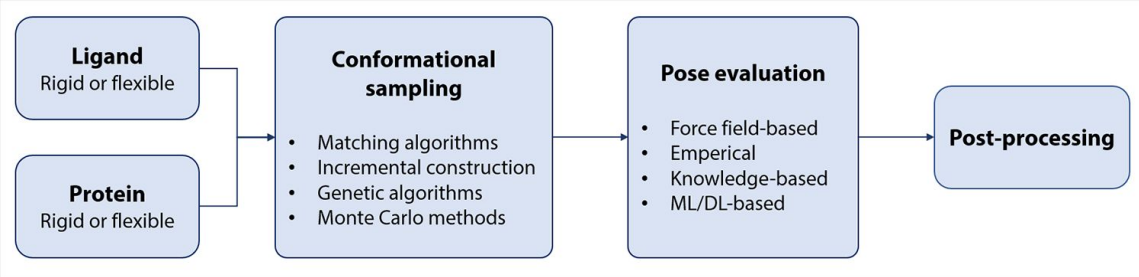
Watch out for biological activity cliffs! Similarity does not imply activity. Three vascular endothelial growth factor receptor 2 (VEGFR2) ligands are shown that represent different similarity−activity relationships.



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.

Protein ligand docking is a commonly used method for structure-based drug design





Docking is like a discotheque: it is all about posing and scoring

Roger Sayle (NextMove Software Limited)

Left: Different strategies to design a ligand in target-based drug discovery: docking (left), building (center), and linking (right). D = H-bond donor, A = H-bond acceptor, H1, H2 = hydrophobic regions of the protein.Applied chemoinformatics: achievements and future opportunities. (Wiley-VCH, 2018), p180. Right: TeachOpenCADD, T015, Protein-ligand docking

The principle of molecular docking

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.

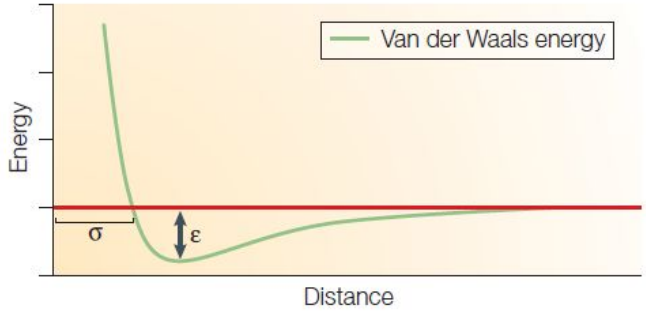
$$E_{coul}(r) = \sum_{i=1}^{N_A} \sum_{j=1}^{N_B} \frac{q_i q_j}{4\pi\varepsilon_0 r_{ij}}$$

Coulombic interactions (electrostatic interactions between electric charges)

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

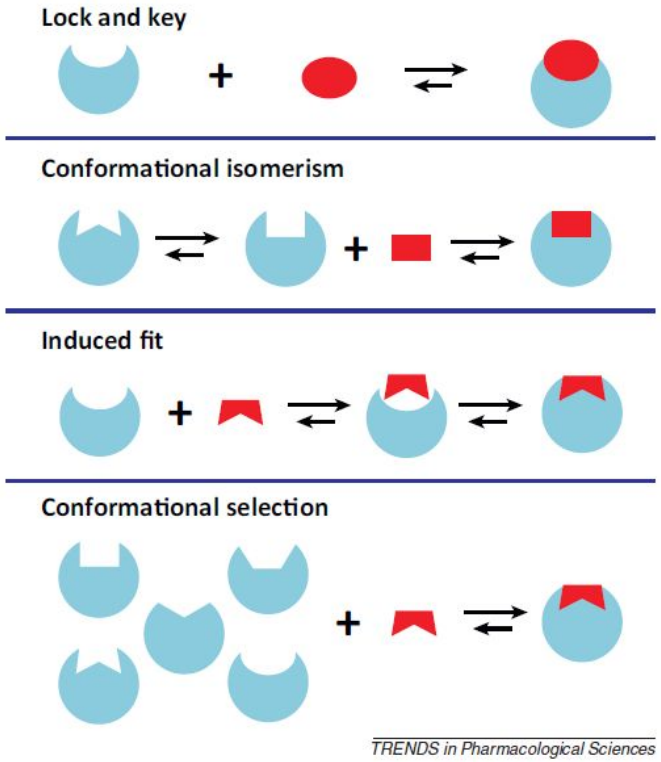
Lennard–Jones 12–6 function (intermolecular interactions without charge)

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

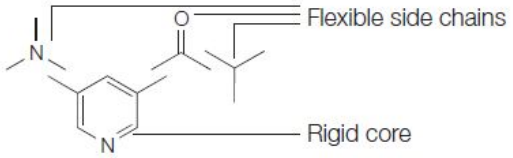


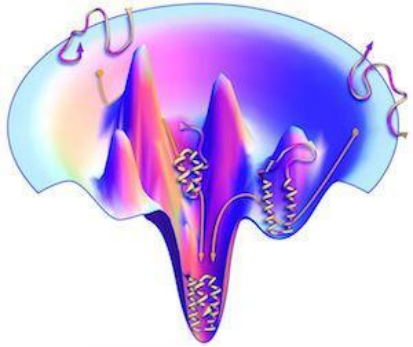
Kitchen, Douglas B., Hélène Decornez, John R. Furr, und Jürgen Bajorath. „Docking and Scoring in Virtual Screening for Drug Discovery: Methods and Applications". *Nature Reviews Drug Discovery* 3, Nr. 11 (November 2004): 935–49. https://doi.org/10.1038/nrd1549.

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.

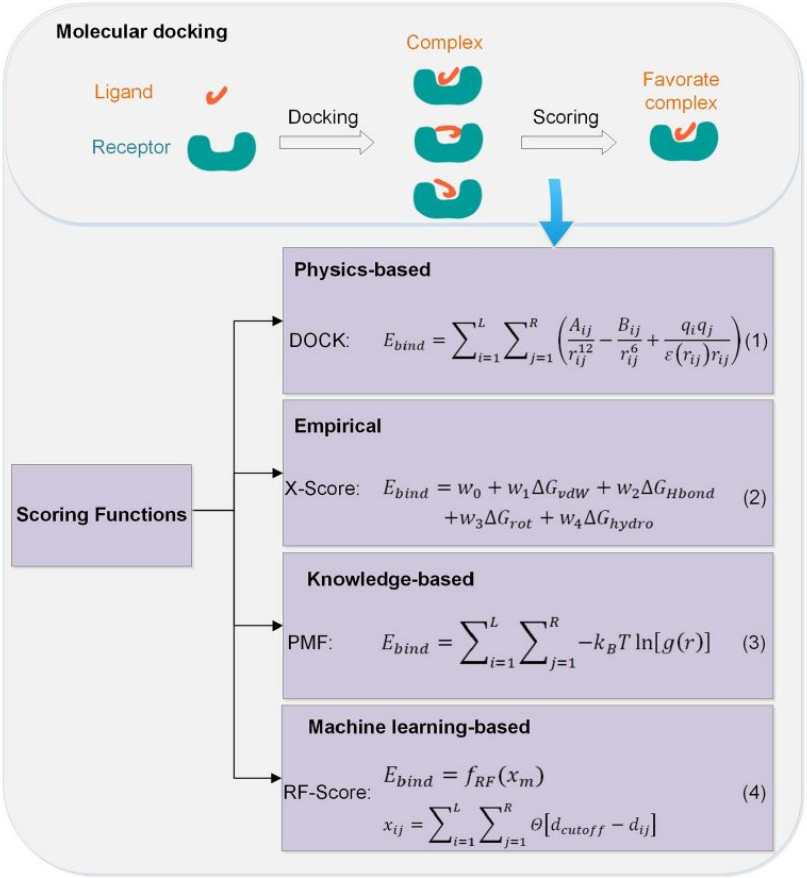




Methods to deal with ligand and protein flexibility

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

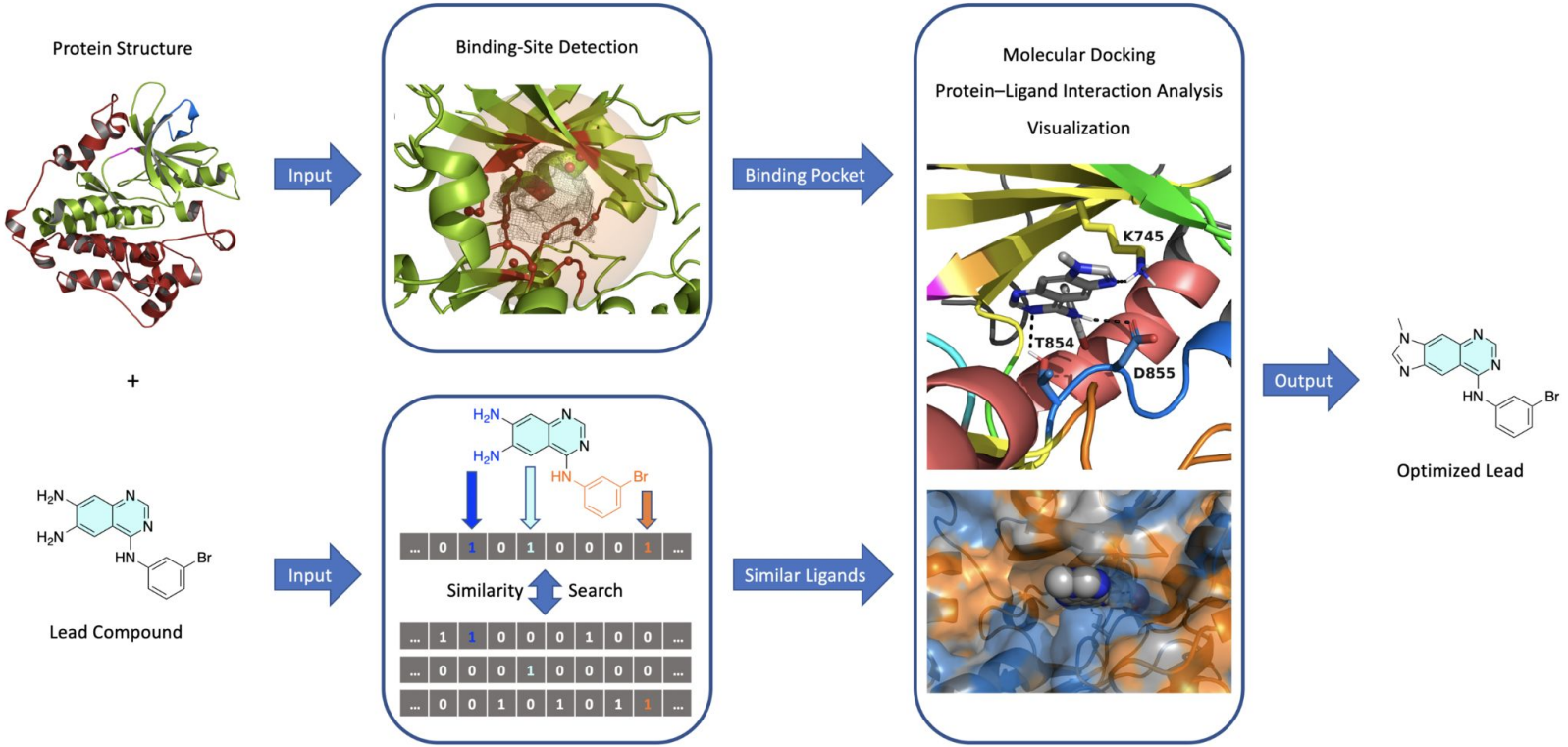
Four types of scoring functions



- **Physics-based (force-field based) scoring:** calculating the energy of individual interactions with force fields.
- **Empirical scoring**: use coefficients to estimate the total energy. The coefficients are estimated from regression analysis of known protein-ligand complexes.
- **Knowledge-based scoring:** integrate results from solved protein-ligand structures, which contains atom-atom contact frequencies and distances. Poses score higher if they show contact characteristics that are often observed in the statistical analysis.
- **Machine learning-based scoring**: use fingerprints or graph. It is usually used in post-processing for rescoring to improve the initial docking.

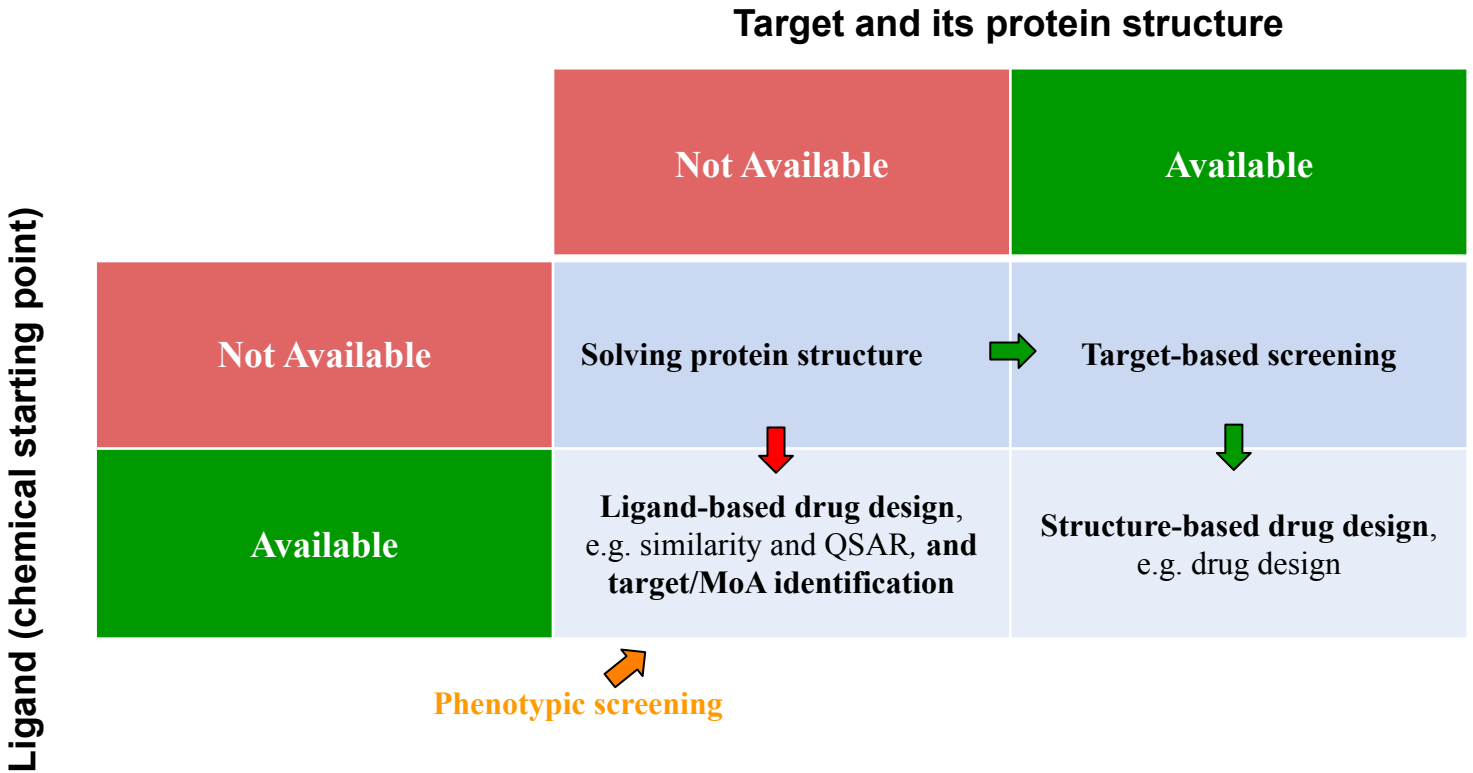
Li, Jin, Ailing Fu, und Le Zhang. „An Overview of Scoring Functions Used for Protein–Ligand Interactions in Molecular Docking“. *Interdisciplinary Sciences: Computational Life Sciences* 11, Nr. 2 (1. Juni 2019): 320–28. https://doi.org/10.1007/s12539-019-00327-w.

A general architecture of analysis pipeline for structure-based virtual screening



Ariamajd, Vogel, Volkamer, Sydow, Taylor, TeachOpenCADD Taltorial T018: Automated pipeline for lead optimization

Structure-based and ligand-based drug design

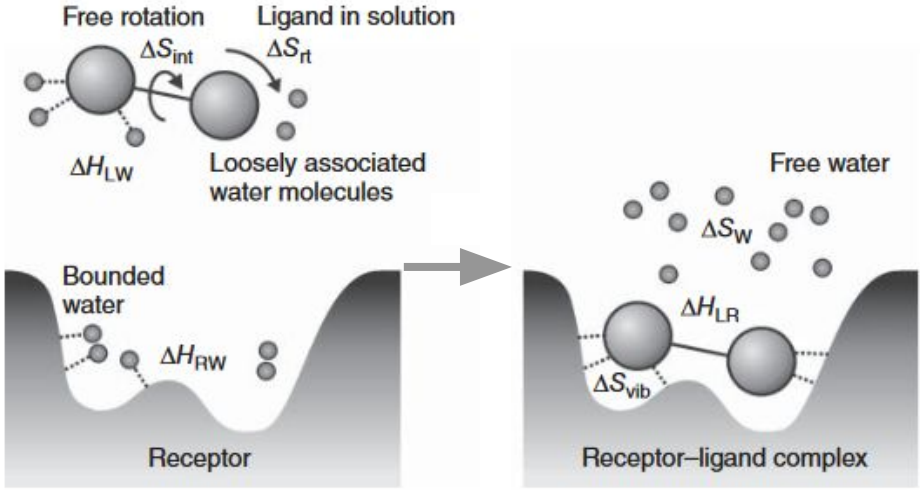


QSAR= quantitative structure activity relationship; MoA= mechanism of action, or mode of action

Thermodynamics and kinetics of ligand-target binding

The **thermodynamics view** of binding: Both enthalpy (heat transfer) and entropy (disorder) contribute to the binding energy (ΔG=ΔH-TΔS).

The **dynamics view** of binding: The *rate* of binding is called affinity, often expressed in K_d (the *Dissociation constant*), written as K_i for inhibitors.



$$\mathrm{L} + \mathrm{R} \underset{k_2}{\overset{k_1}{\rightleftharpoons}} \mathrm{LR} \qquad \mathrm{E} + \mathrm{S} \underset{k_r}{\overset{k_f}{\rightleftharpoons}} \mathrm{ES} \xrightarrow{k_{cat}} \mathrm{E} + \mathrm{P}$$

$$K_D \equiv k_2/k_1 \qquad K_D \equiv \frac{k_r}{k_f} \qquad V_{max} \equiv k_{cat}[E]_0$$

$$[LR] = \frac{[R_{total}][L]}{[L] + K_D} \qquad v = \frac{V_{max}[S]}{K_D + [S]}$$

The dose-response curve and IC50: The Hill function and *in vitro* pharmacology

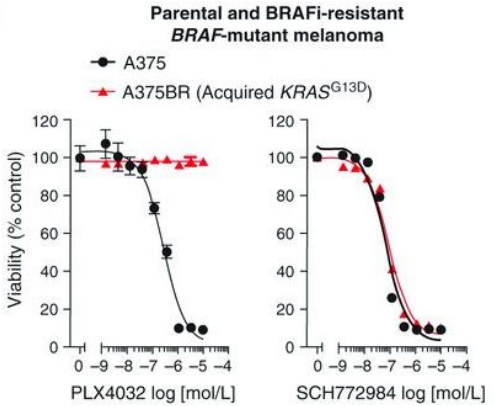
- The Hill function is one of the mostly useful non-linear functions to model biological systems.
- In its general form, H_{max} indicates the maximal value to which the function is asymptotic, *n* is the shape parameter (known as the Hill's coefficient), and *k* is the reflection point, often abbreviated 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)*.

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

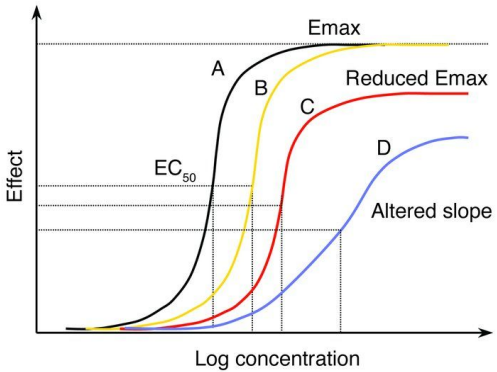
The general form of the Hill function

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

Modelling the dose-dependent effect



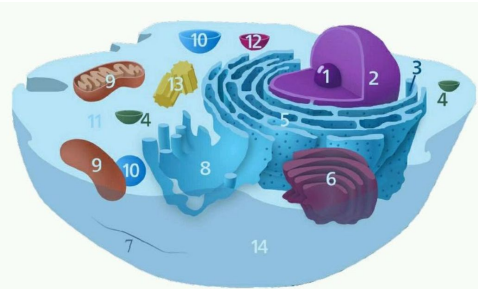
Morris *et al.* Cancer Discov; 3(7); 742–50. ©2013 AACR.

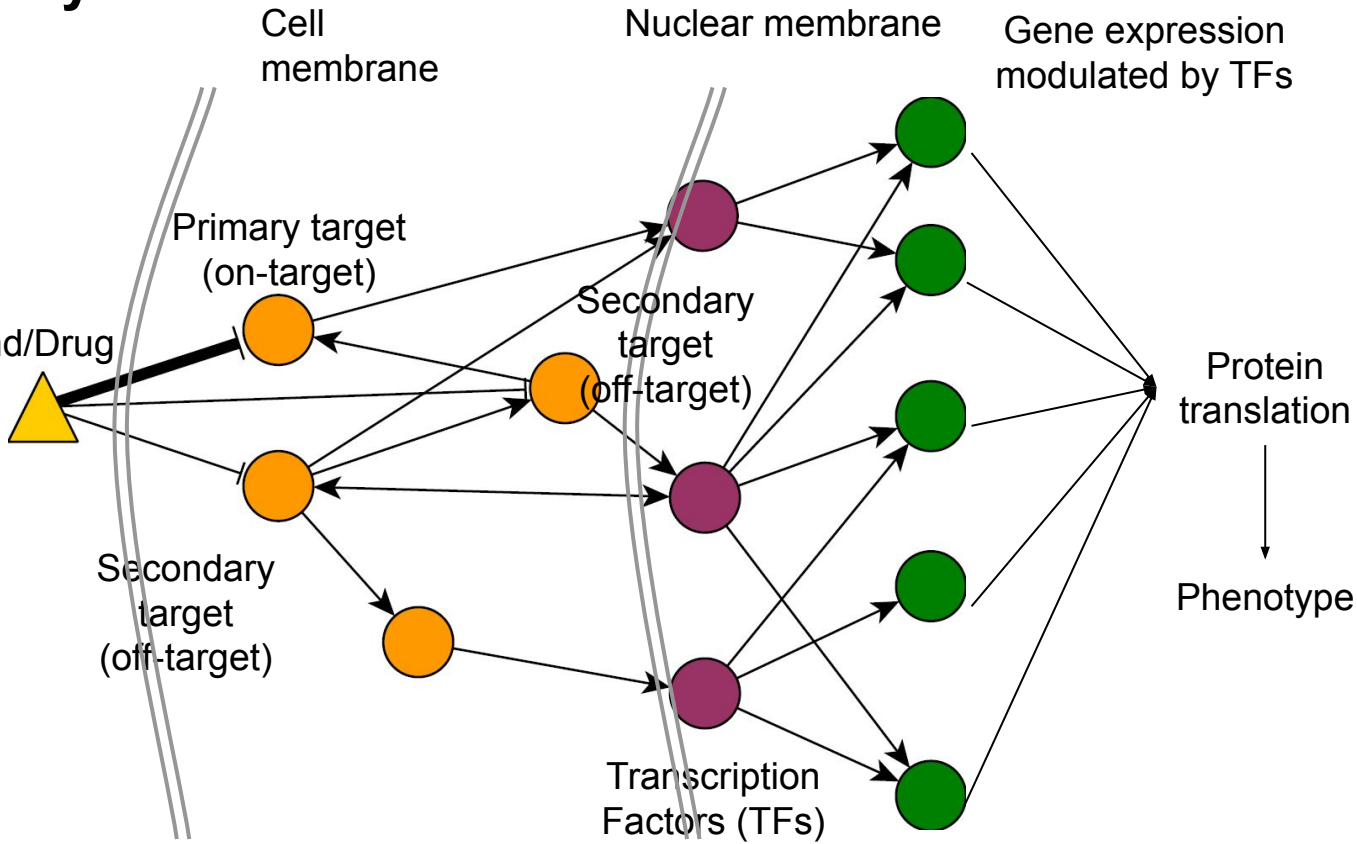


White. *J Clin Invest.* 2004;113(8):1084-1092. https://doi.org/10.1172/JCI21682.

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

Biological networks interact with drugs and manifest its efficacy and safety





Reaction Rate Equations: a compartment/ODE model of biological chemical reaction

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

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

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

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

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

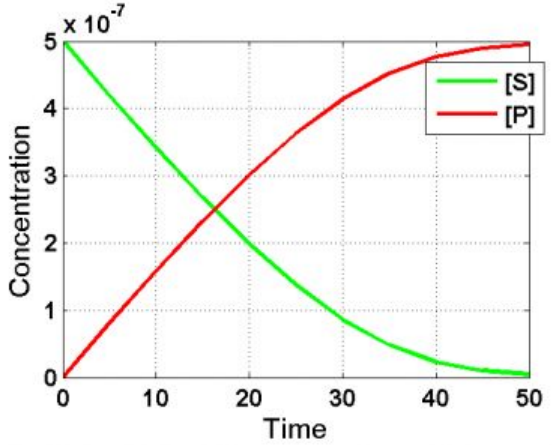
$$K_D \equiv \frac{k_r}{k_f}$$

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

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

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

***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*.

Simulation of biological networks with ordinary differential expression

Given the reaction

$$\mathrm{S} + \mathrm{E} \underset{k_2}{\overset{k_1}{\rightleftharpoons}} \mathrm{C} \xrightarrow{k_3} \mathrm{P} + \mathrm{E}$$

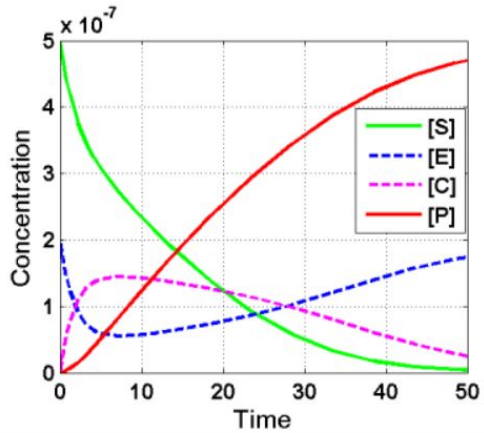
Given the initial values and rate constants

- $S(0) = 5e^{-7}$
- $E(0) = 2e^{-7}$
- $C(0) = P(0) = 0$
- $k_1 = 1e^{6}$
- $k_2 = 1e^{-4}$
- $k_3 = 0.1$

According to the law of mass action

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

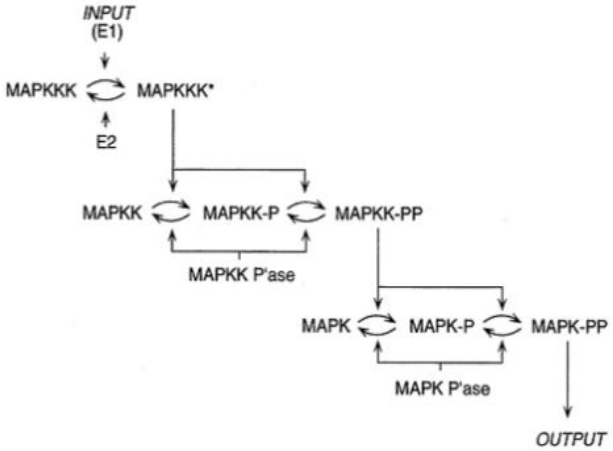
It is possible to simulate the concentration changes by time *deterministically*.

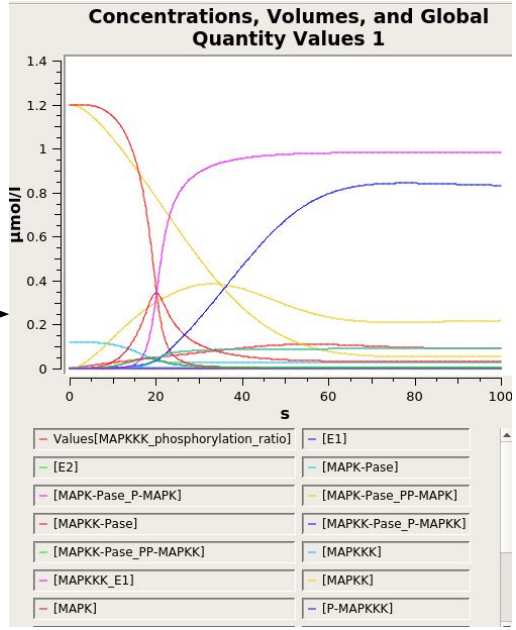


See Systems Engineering Wiki (tue.nl) for MATLAB/COPASI codes and *Stochastic Modelling for Systems Biology* by Darren J. Wilkinson

Simulating behavior of complex ODE systems with COPASI

- COPASI, freely available at http://COPASI.org/, supports both **ordinary differential equation (ODE)** based simulation as well as stochastic kinetic simulation.
- Such tools are important for detailed analysis of enzymatic reactions, for instance in the presence of drugs and/or disease-relevant mutation.

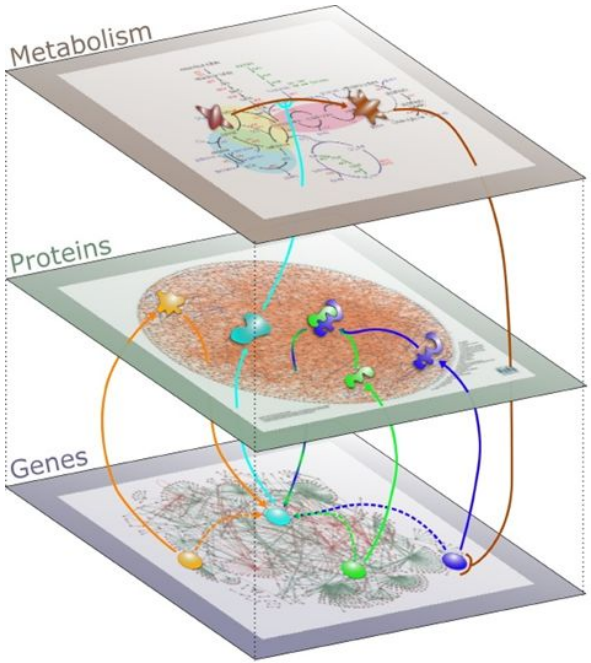




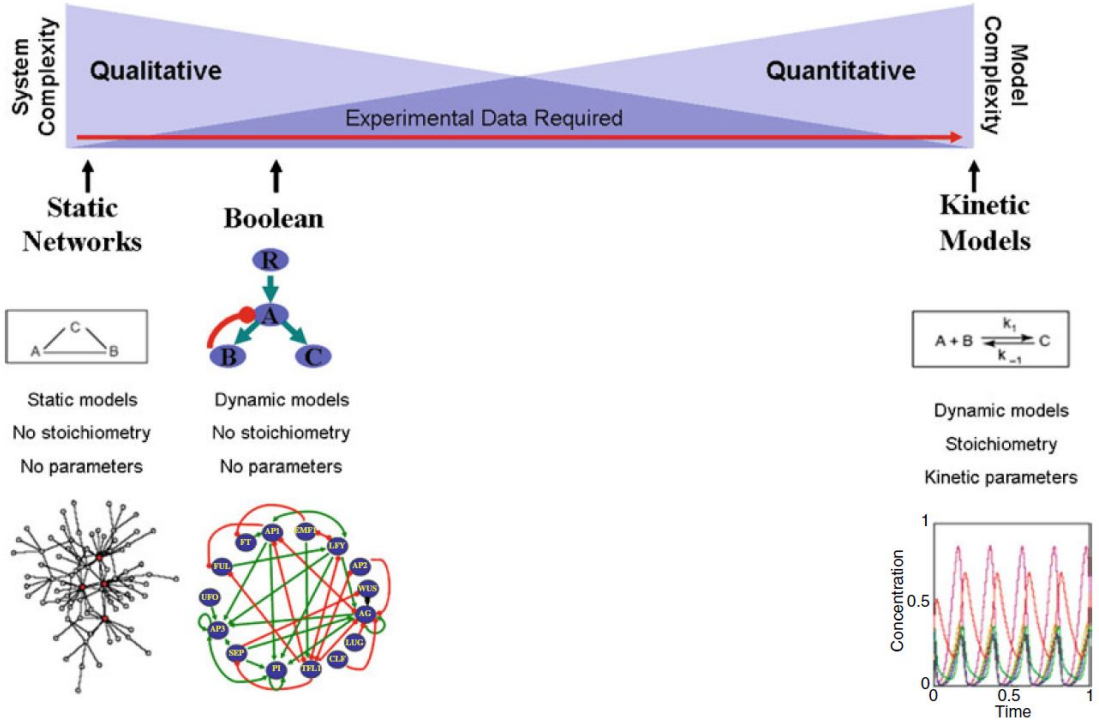
ODE-based simulation of dynamics

Figure: Huang and Ferrell, PNAS, 2006. 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)

Modelling biological networks



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



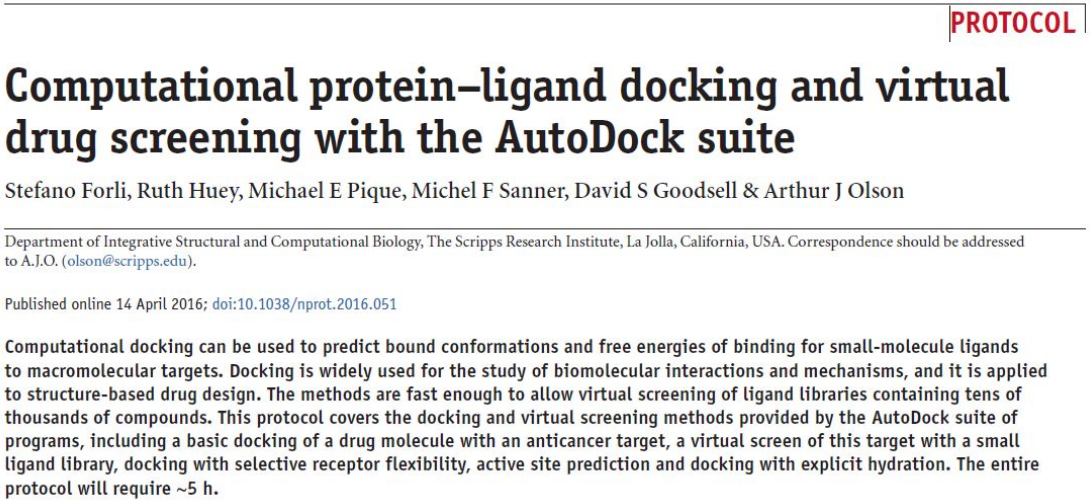
Garg, Abhishek, Kartik Mohanram, Giovanni De Micheli, and Ioannis Xenarios. 2012. "Implicit Methods for Qualitative Modeling of Gene Regulatory Networks." In *Gene Regulatory Networks: Methods and Protocols*, edited by Bart Deplancke and Nele Gheldof, 397–443. Methods in Molecular Biology. Totowa, NJ: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

Offline activities

- Read selected pages of *Computational Methods in Drug Discovery* by Sliwoski *et al*. Please submit your results by using this Google Form **by December 1st**.

Backup slides

Resources for 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.

Caution: Binding predicted by docking should always be challenged and verified by experimental testing! Docking scores seldomly correlate with binding affinity.

- 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.

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.

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]

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

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

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

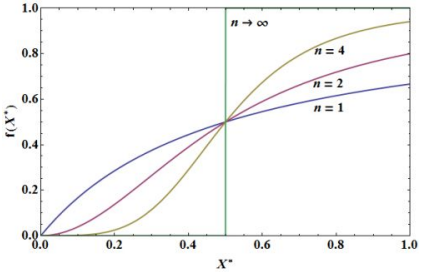
$$[LR] = \frac{k_1[L][R_{total}]}{k_1[L] + k_2}$$

$K_D \equiv k_2/k_1$

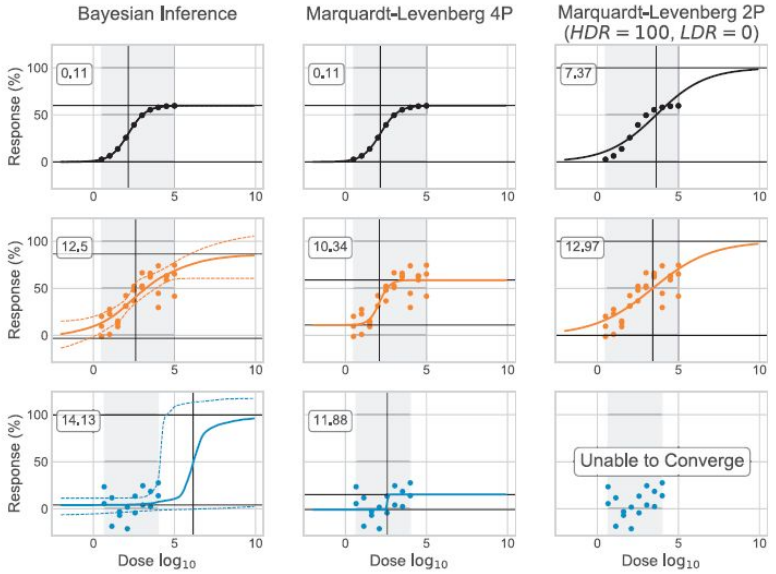
$$[LR] = \frac{[R_{total}][L]}{[L] + K_D}$$

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.



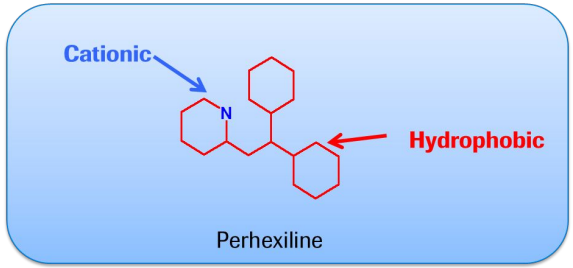
From the biophysics wiki article by Andreas Piehler



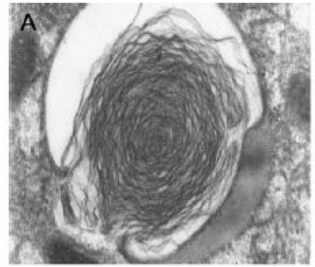
The Bayesian inference approach versus the non-Bayesian Marquardt-Levenberg algorithm for non-linear regression fitting. Labelle, Caroline, Anne Marinier, and Sébastien Lemieux. 2019. "Enhancing the Drug Discovery Process: Bayesian Inference for the Analysis and Comparison of Dose–Response Experiments." *Bioinformatics* 35 (14): i464–73.

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

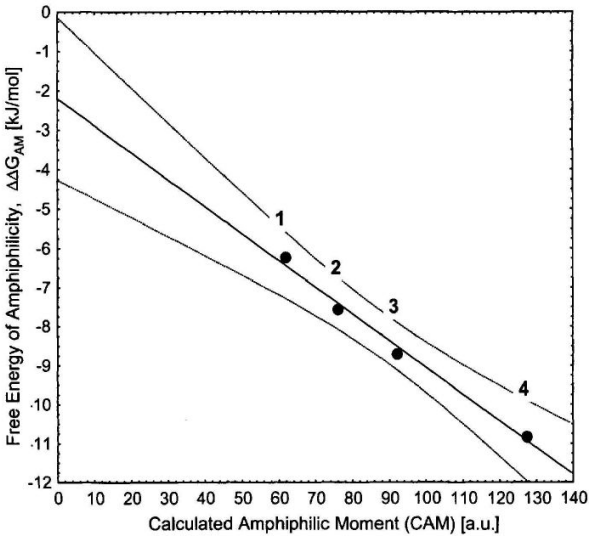
d: distance between the center of gravity of the charged part of a molecule and the hydrophobic/hydrophilic remnant of the molecule

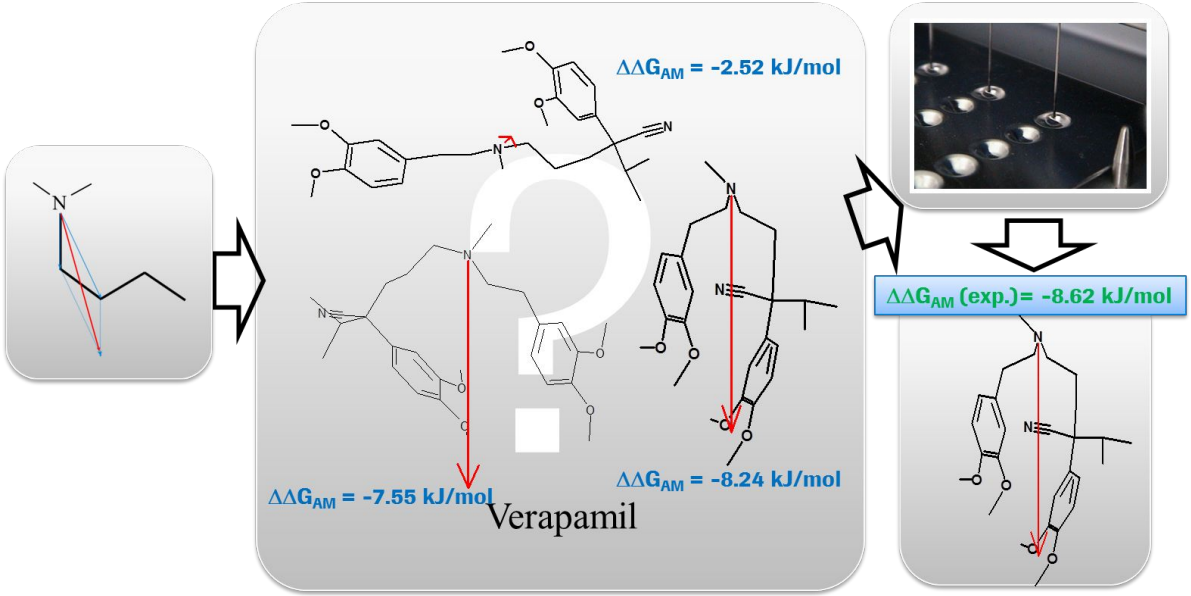
$\vec{\alpha}_i$: the hydrophobic/hydrophilic contribution of atom/fragment *i*

***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)

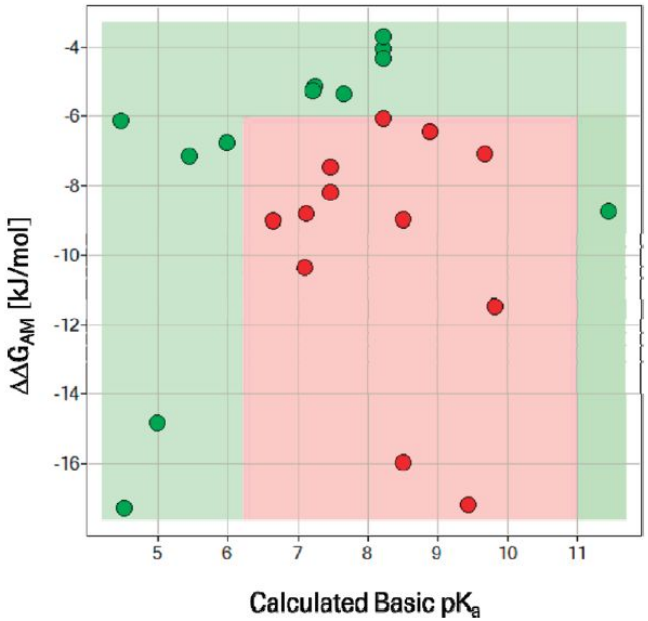




Iterative model building, experimentation, and model refining led to the predictive tool CAFCA

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 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, 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).
- **Safety liabilities caused by physicochemical properties of the drugs may be well predicted by molecular modelling inspired by simple models.**

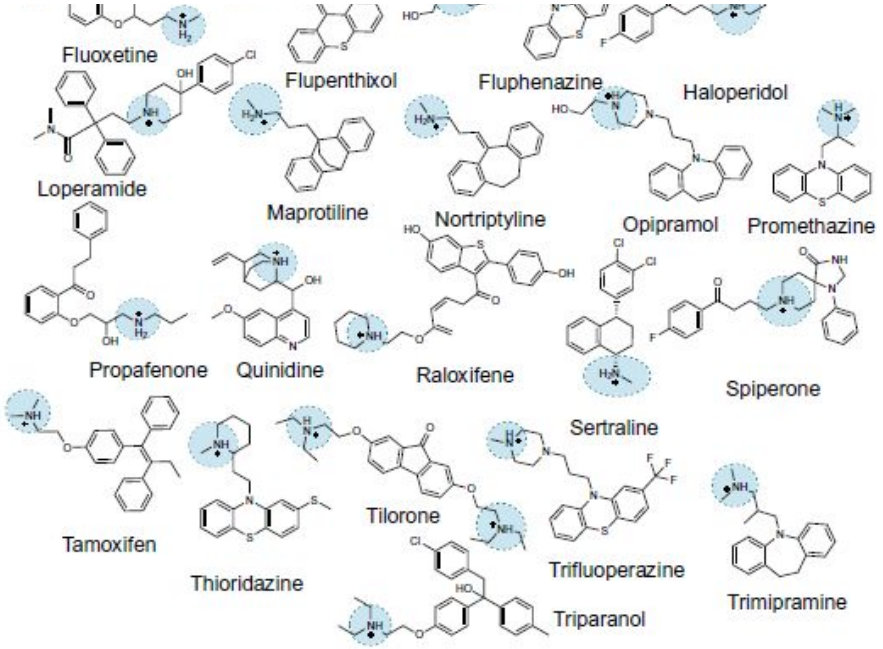


Fig. 1. Representative examples of CADs that are identified in SARS-CoV-2 drug repurposing screens.

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.

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

Given the reaction $\mathrm{A} + \mathrm{B} \underset{k_2}{\overset{k_1}{\rightleftharpoons}} \mathrm{C} + \mathrm{D}$ 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 $\begin{bmatrix} K \\ K \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} K-1 \\ K-1 \\ 1 \\ 1 \end{bmatrix}, \begin{bmatrix} K-2 \\ K-2 \\ 2 \\ 2 \end{bmatrix}, \ldots, \begin{bmatrix} 0 \\ 0 \\ K \\ K \end{bmatrix},$

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.

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:
  1. **initialize** the system with initial conditions
  2. 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.
  3. Let the randomly selected reaction happen and **update** the time.
  4. **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.

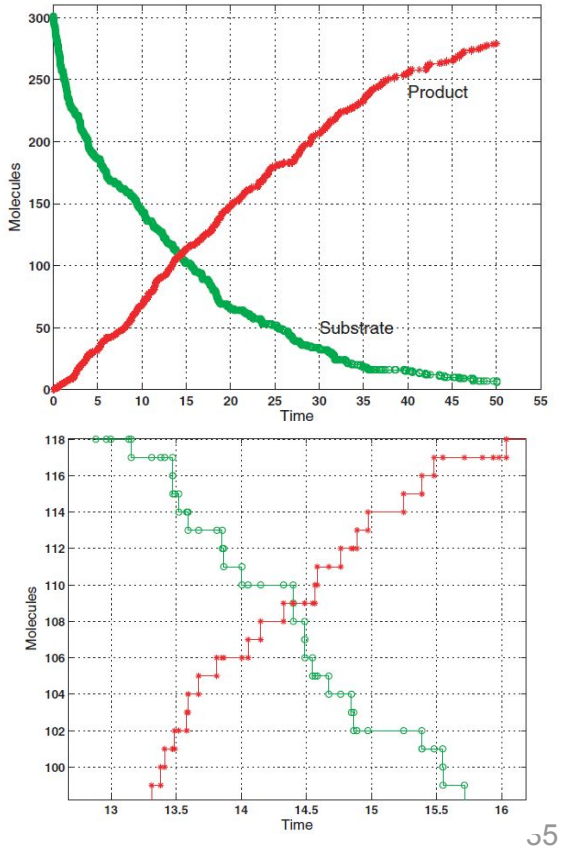
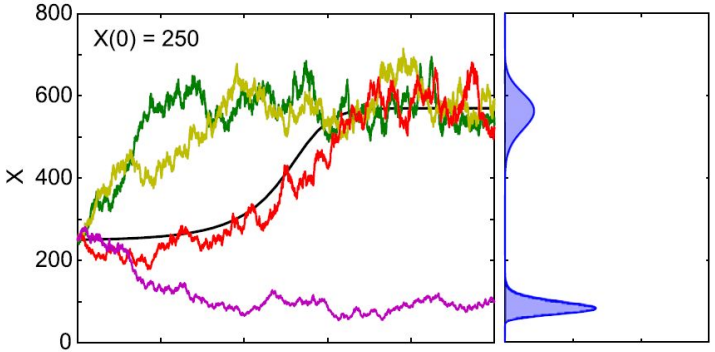
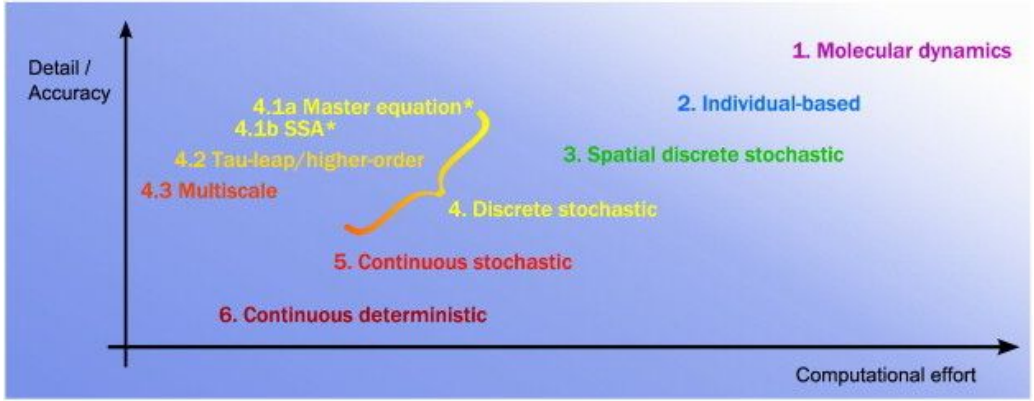


Figure source and further reading: Higham, Desmond J. 2008. "Modeling and Simulating Chemical Reactions." *SIAM Review* 50 (2): 347–68. https://doi.org/10.1137/060666457.

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.

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.

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]
Tau-leap	4	Relatively fast	Approximate; too slow for large systems or frequent/multiscale reactions	[83,113,118]	StochKit [145]
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	Continous; no noise	[156]	
ODE	6	Very fast	Continuous; no noise	[157]	

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.

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.