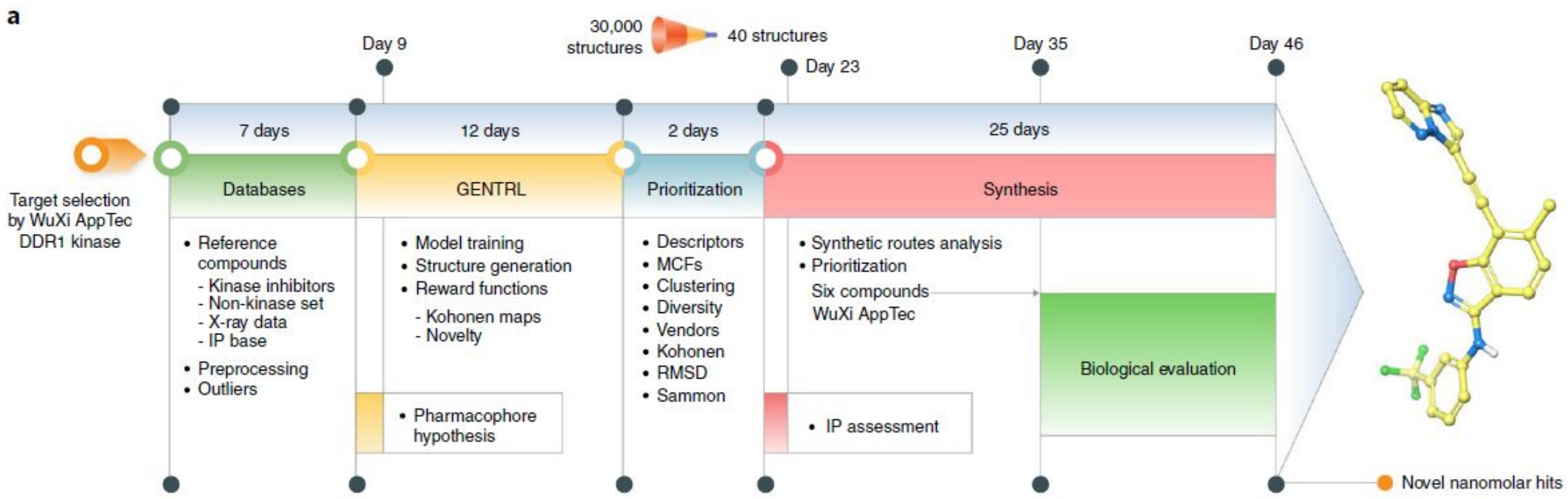
AMIDD Lecture 5: Principles of Molecular modelling



Deep learning enables rapid identification of potent DDR1 kinase inhibitors, Zhavoronkov *et al.*, Nature Biotechnology, 2019. Source code: https://github.com/insilicomedicine/gentrl

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

Outline

1. The two views of ligand-receptor interaction
2. The Michaelis-Menten model
3. The Hill equation
4. Structure-based molecular modelling, with molecular docking as an example
5. Ligand-based molecular modelling, with QSAR as an example

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]

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

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

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

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

$K_D \equiv k_2/k_1$

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

Four classical classes of mathematical models

Compartment models

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

Kinetics of ligand-target interaction

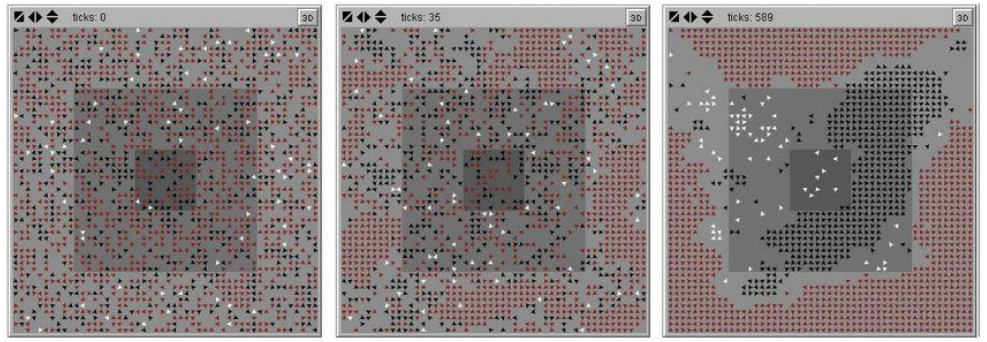
$$\frac{dx}{dt} = \alpha x - \beta xy,$$
$$\frac{dy}{dt} = -\gamma y + \delta xy,$$

The Lotka-Volterra equations modelling predator-prey relationships.

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

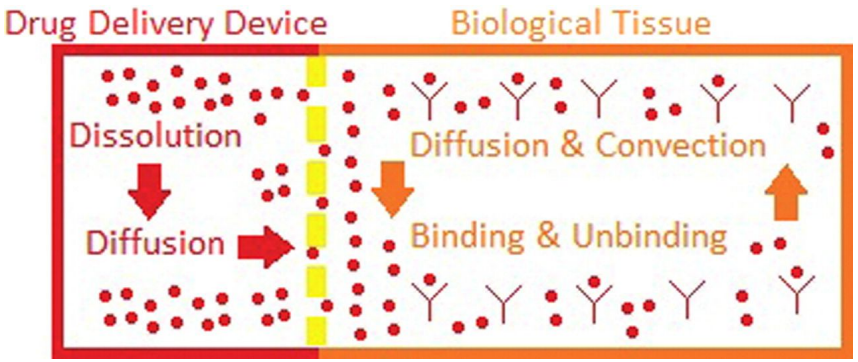
The SIR (S=susceptible, I=infectious, R=removed) model of epidemiology

Particle models



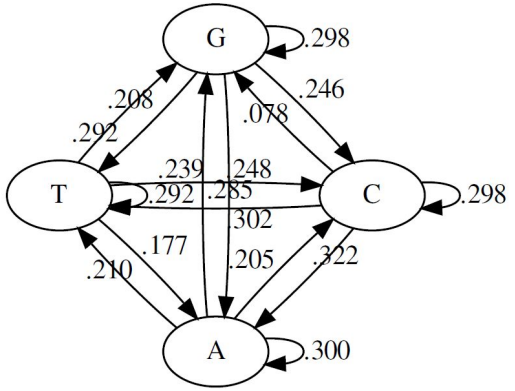
A Study on Socio-spatial Segregation Models Based on Multi-agent Systems by Quadros *et al*. (2012). 10.1109/BWSS.2012.14.

Transport models



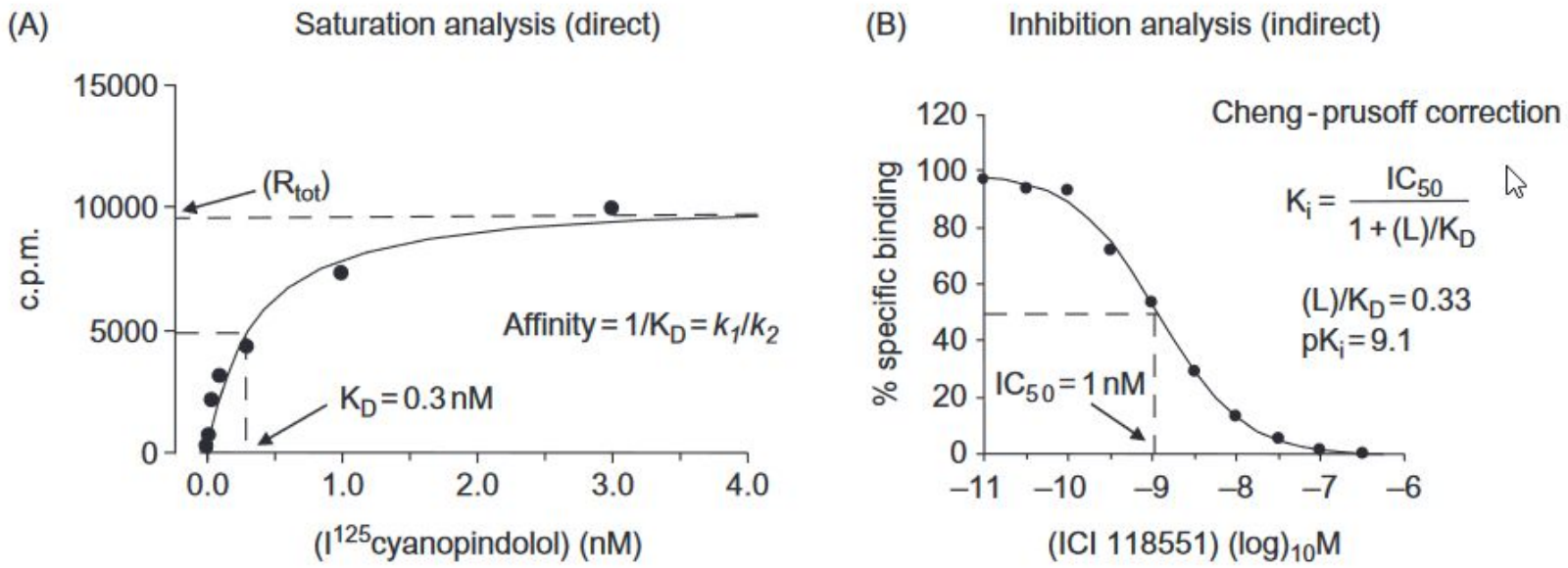
McGinty, Sean, and Giuseppe Pontrelli. 2015. "*A General Model of Coupled Drug Release and Tissue Absorption for Drug Delivery Devices*." Journal of Controlled Release 217 (November): 327–36.

Finite state models



A finite-state Markov chain modelling DNA sequences

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?

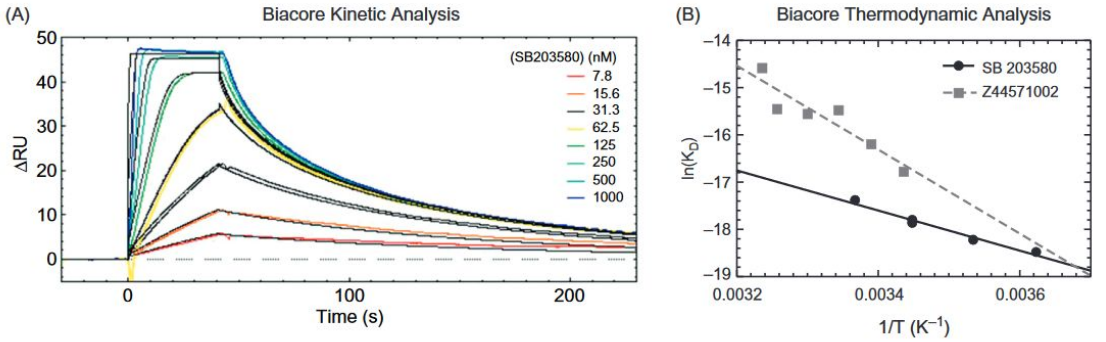
The biophysical (thermodynamic) view of binding affinity: enthalpy and entropy

Gibbs free energy

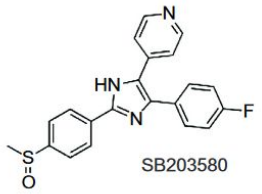
$$\Delta G = \Delta H - T\Delta S$$

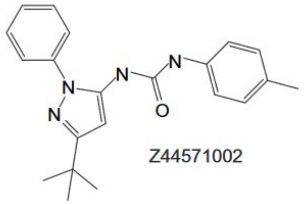
Van't Hoff Equation

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



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.*

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

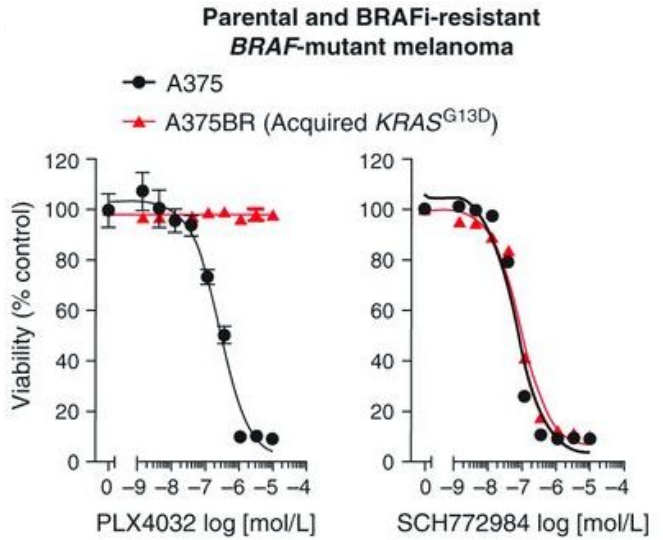
- The Hill function is one of the mostly useful non-linear functions to model biological systems.
- In its general form, H_{max} indicates the maximal value to which the function is asymptotic, *n* is the shape parameter (known as the Hill's coefficient), and *k* is the reflection point, often abbreviated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

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

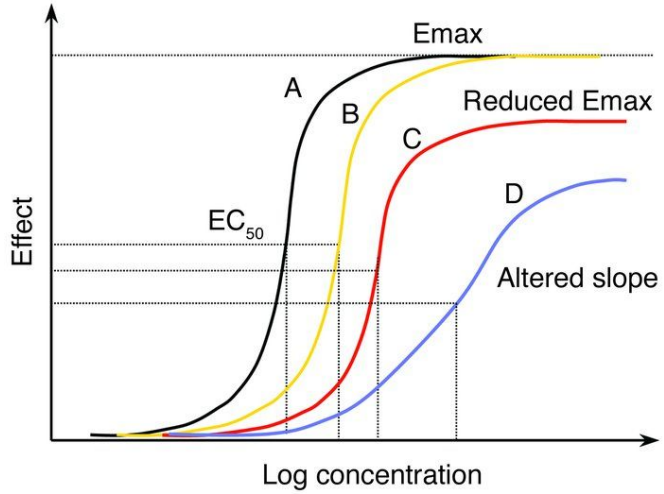
General form of the Hill function

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

Modelling dose-dependent effect



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

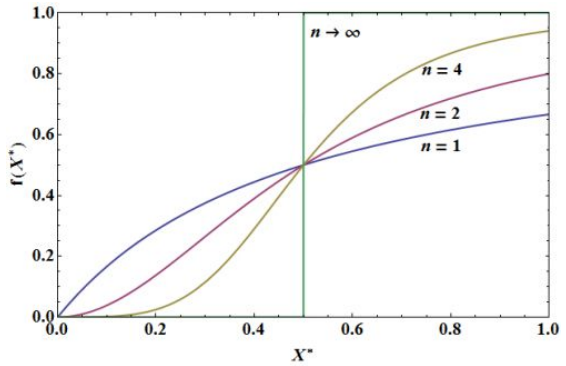


White. *J Clin Invest.* 2004;113(8):1084-1092. https://doi.org/10.1172/JCI21682.

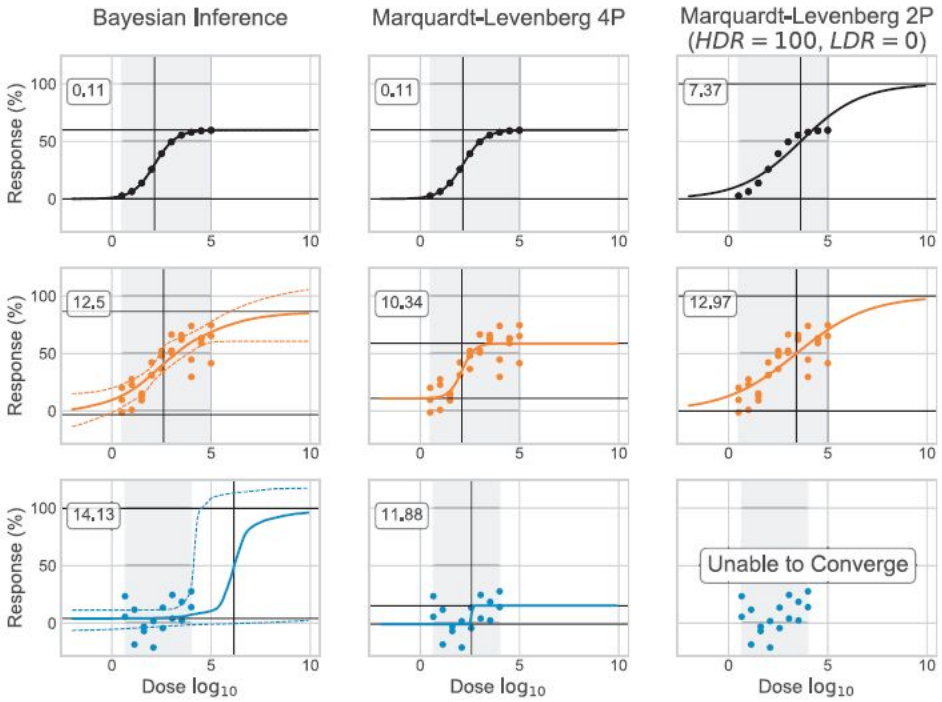
Suppose it is an antiviral drug, compared with curve B, what does curve A, C, and D suggest?

More about the Hill function and dose-response curves

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



From the biophysics wiki article by Andreas Piehler



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

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

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

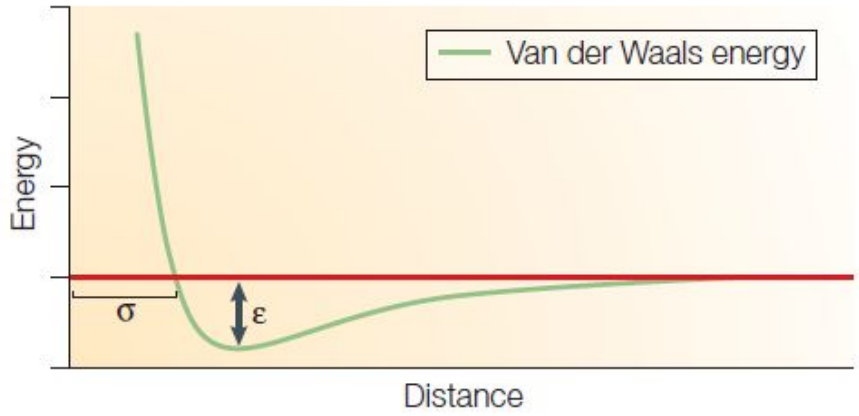
$$E_{coul}(r) = \sum_{i=1}^{N_A} \sum_{j=1}^{N_B} \frac{q_i q_j}{4\pi\varepsilon_0 r_{ij}}$$

Coulombic interactions

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

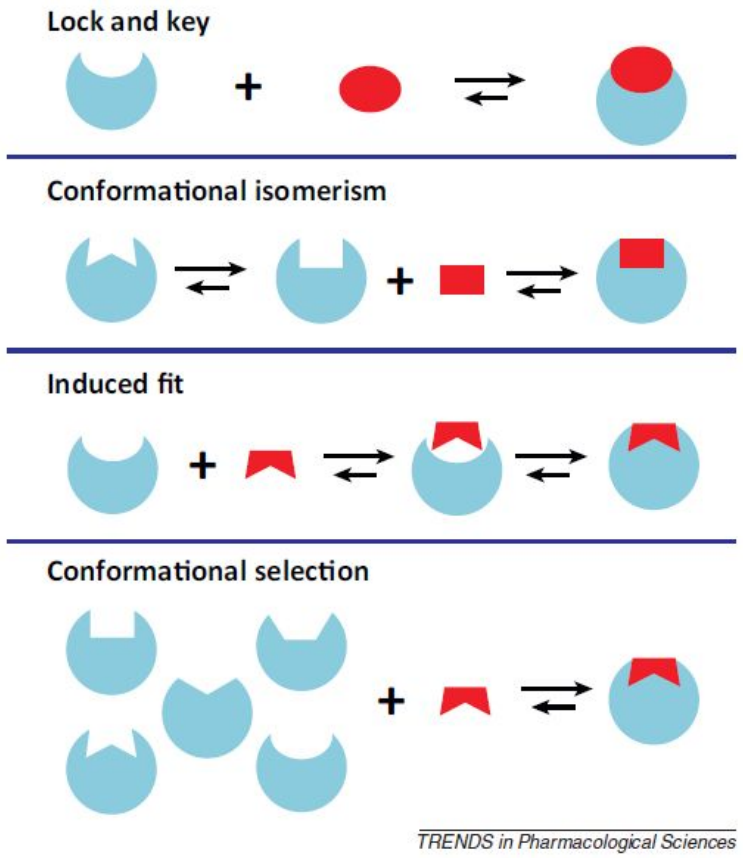
Lennard–Jones 12–6 function

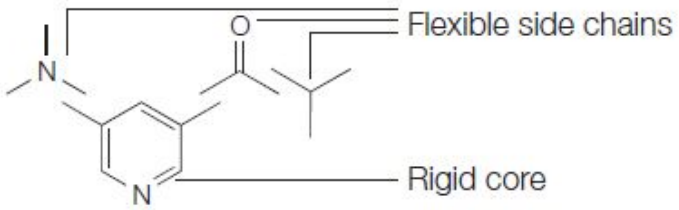
- ε is the **well depth** of the potential
- σ is the **collision diameter** of the respective atoms *i* and *j*.

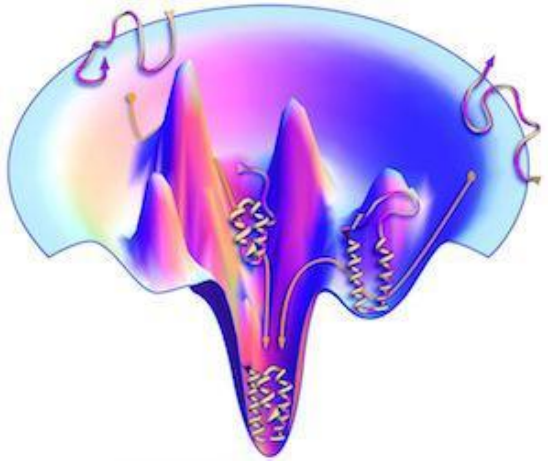


Kitchen, Douglas B., Hélène Decornez, John R. Furr, und Jürgen Bajorath. „Docking and Scoring in Virtual Screening for Drug Discovery: Methods and Applications". *Nature Reviews Drug Discovery* 3, Nr. 11 (November 2004): 935–49. https://doi.org/10.1038/nrd1549.

Posing: dealing with flexibility of ligand and of protein





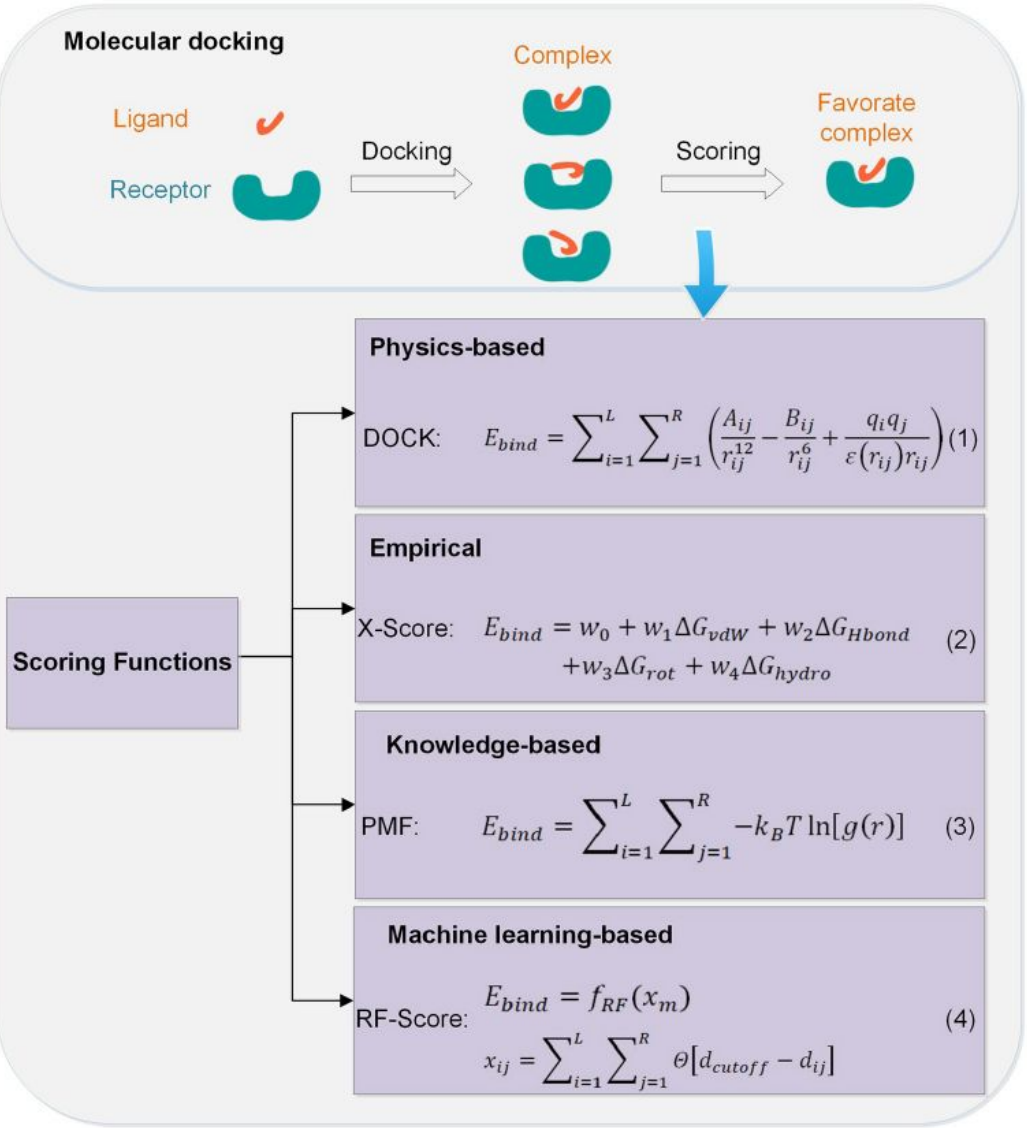


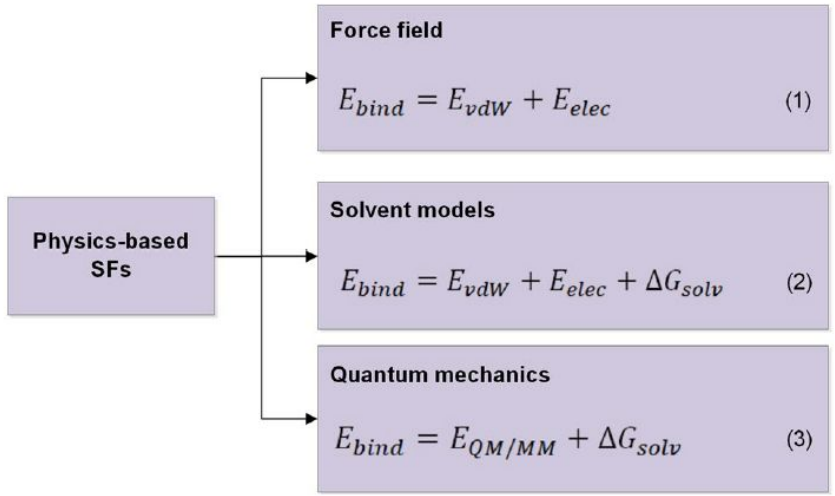
Methods to deal with ligand and protein flexibility

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

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.

Types of scoring functions

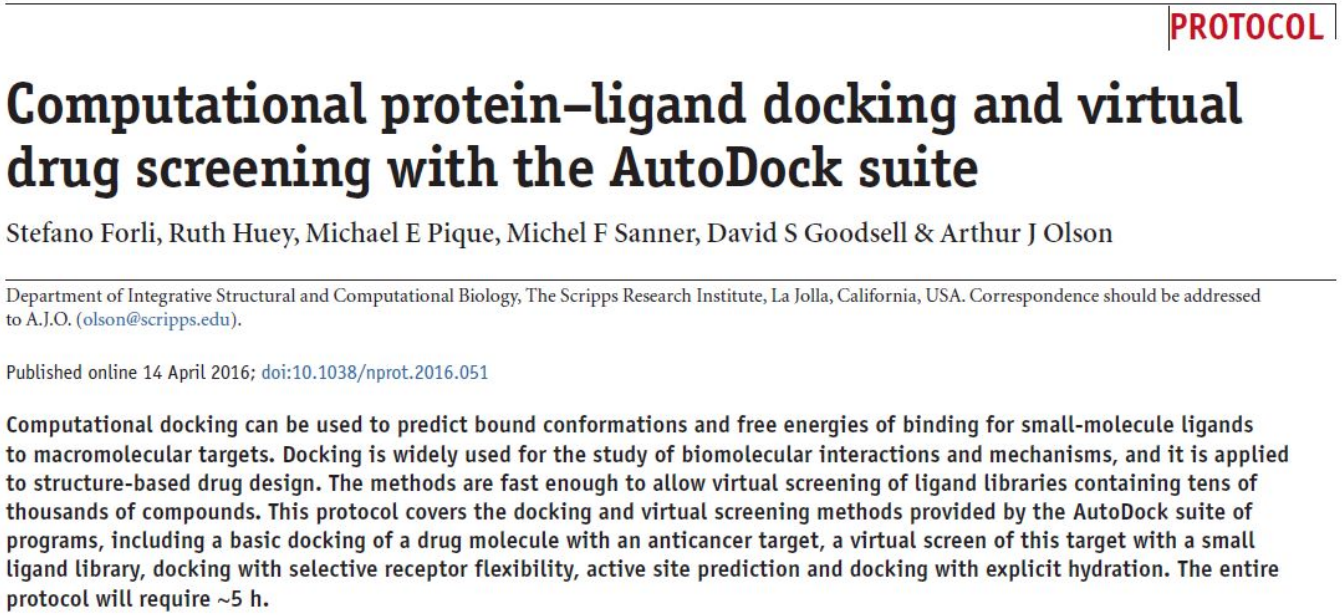




- Empirical scoring functions estimate the binding affinity of a complex by **summing up the important energetic factors for protein–ligand binding**, such as hydrogen bonds, hydrophobic effects, steric clashes, etc. It relies on training set and regression analysis.
- Knowledge-based scoring functions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

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

Interested in learning more about molecular modelling?



PROTOCOL

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

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

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

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

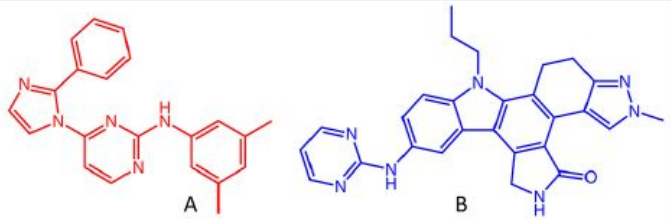
Computational docking can be used to predict bound conformations and free energies of binding for small-molecule ligands to macromolecular targets. Docking is widely used for the study of biomolecular interactions and mechanisms, and it is applied to structure-based drug design. The methods are fast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

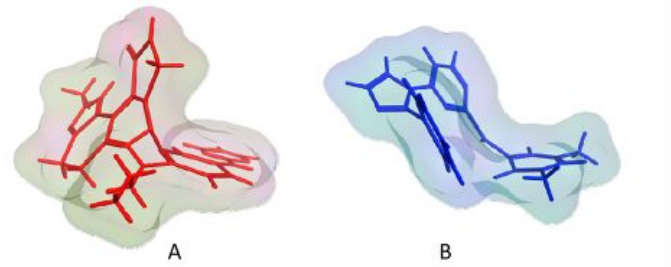
- Try docking yourself by following this protocol: Forli, Stefano, Ruth Huey, Michael E. Pique, Michel F. Sanner, David S. Goodsell, und Arthur J. Olson. „Computational Protein–Ligand Docking and Virtual Drug Screening with the AutoDock Suite“. *Nature Protocols* 11, Nr. 5 (Mai 2016): 905–19. https://doi.org/10.1038/nprot.2016.051.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

Molecular similarity including the Tanimoto (Jaccard) Index

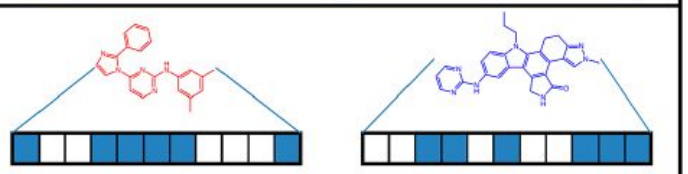
Chemical similarity	
Molecular similarity	
2D similarity	
3D similarity	
Biological similarity	
Global similarity	
Local similarity	

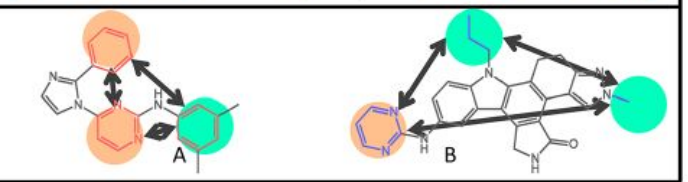
	Mol. weight	LogP	Rotatable bonds	Aromatic rings	Heavy atoms
A	341.4	5.23	4	4	26
B	463.5	4.43	4	5	35



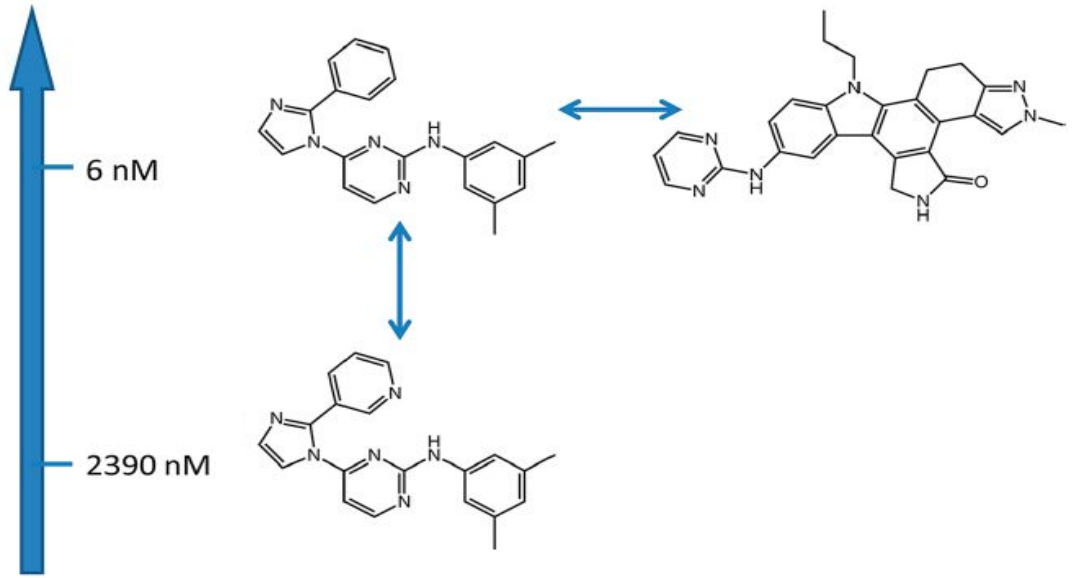


	Vascular endothelial growth factor receptor 2	Tyrosine-protein kinase TIE-2
A	active	inactive
B	active	active





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. https://doi.org/10.1021/jm401411z.



A recent expansion of the similarity principle to other levels of biology: Duran-Frigola, Miquel, Eduardo Pauls, Oriol Guitart-Pla, Martino Bertoni, Víctor Alcalde, David Amat, Teresa Juan-Blanco, and Patrick Aloy. 2020. “Extending the Small-Molecule Similarity Principle to All Levels of Biology with the Chemical Checker.” Nature Biotechnology, May, 1–10.

Quantitative Structure-Activity Relationships (QSARs)

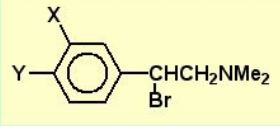
QSAR is a statistical modelling of correlation between biological activity and physicochemical properties. It is among the earliest subjects that used machine learning and pattern recognition in drug discovery.

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.



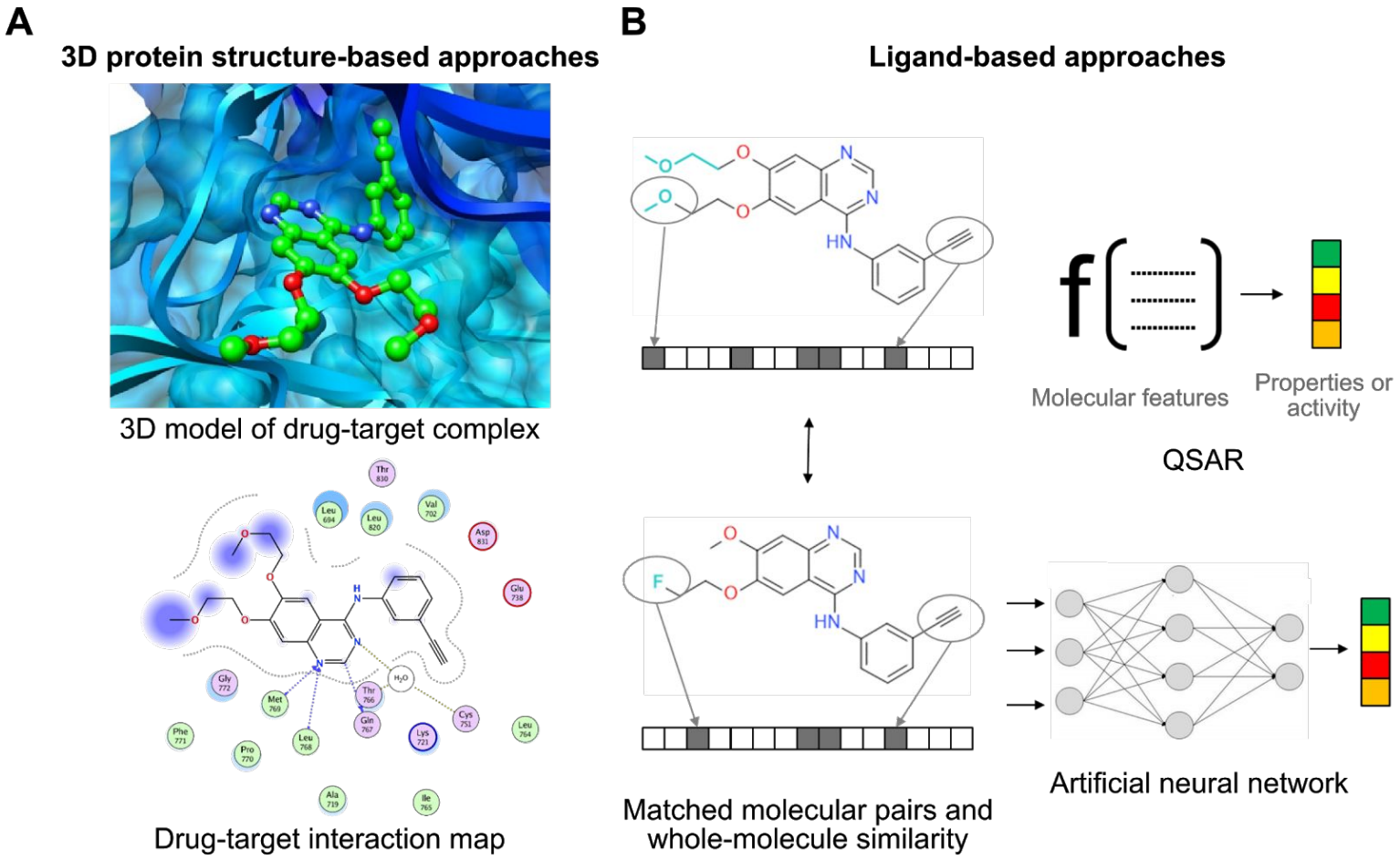
meta	*para*	*meta-*					*para-*					log 1/C	log 1/C
(X)	(Y)	F	Cl	Br	I	Me	F	Cl	Br	I	Me	obsd.	calc.a)
H	H											7.46	7.82
H	F						1					8.16	8.16
H	Cl							1				8.68	8.59
H	Br								1			8.89	8.84
H	I									1		9.25	9.25
H	Me										1	9.30	9.08
F	H	1										7.52	7.52
Cl	H		1									8.16	8.03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

Multivariate regression analysis

$$\log(1/ED_{50}) = -0.301[m\text{–F}] + 0.27[m\text{–Cl}] + 0.434[m\text{–Br}] + 0.579[m\text{–I}] + 0.454[m\text{–Me}] + 0.340[p\text{–F}] + 0.768[p\text{–Cl}] + 1.020[p\text{–Br}] + 1.429[p\text{–I}] + 1.256[p\text{–Me}] + 7.821$$

$$n = 22,\ r^2 = 0.94,\ s = 0.194,\ F = 17.0$$

Summary and Q&A



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

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

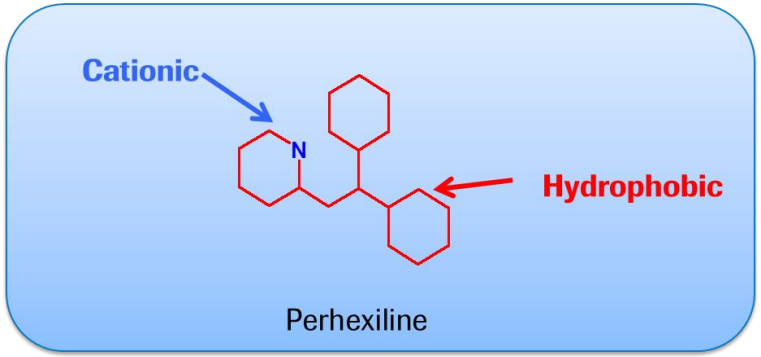
Summary and Q&A

More about the the Free-Wilson analysis

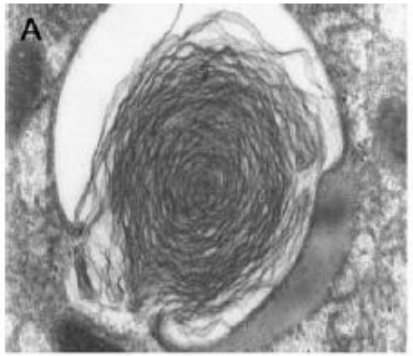
- *A Mathematical Contribution to Structure-Activity Studies* by Spencer M. Free and James W. Wilson, Journal of Medicinal Chemistry, 1964, and reviewed by Kubinyi, 1988.
- A Python implementation on GitHub, and a blog post going through examples, is shared by Pat Walters.
- Free-Wilson nonadditivity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

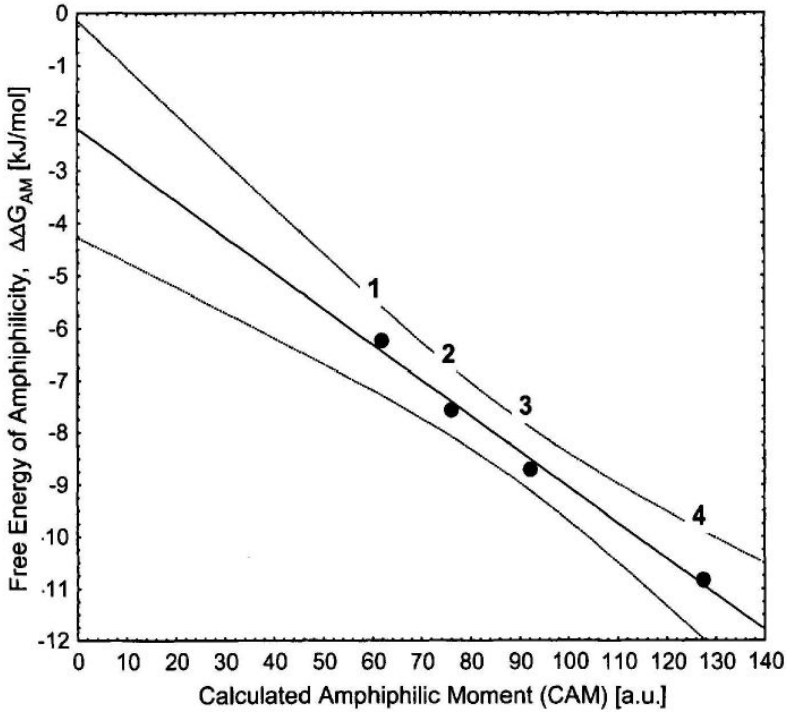
d: distance between the center of gravity of the charged part of a molecule and the hydrophobic/hydrophilic remnant of the molecule

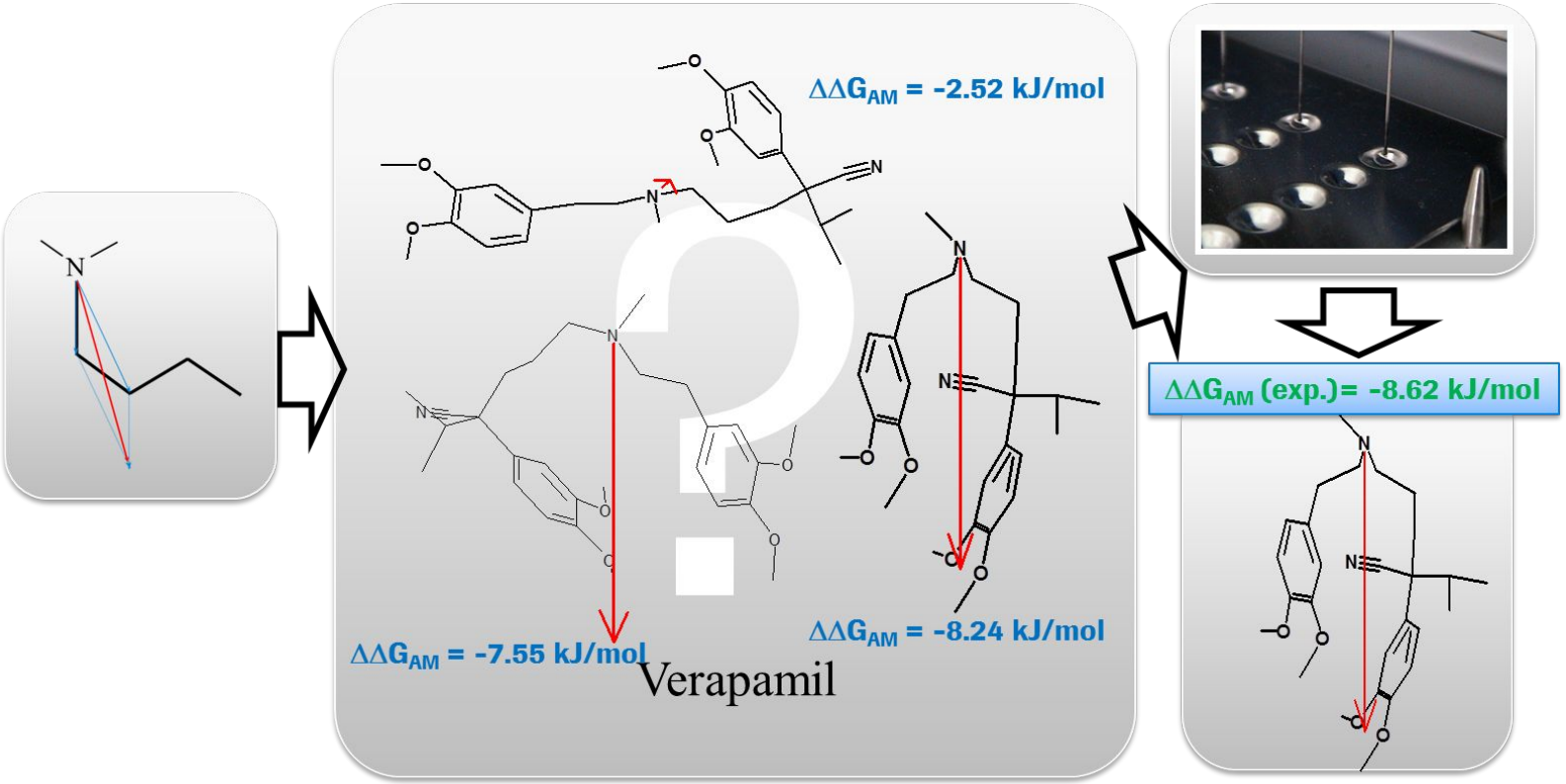
$\vec{\alpha}_i$: the hydrophobic/hydrophilic contribution of atom/fragment *i*

***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)

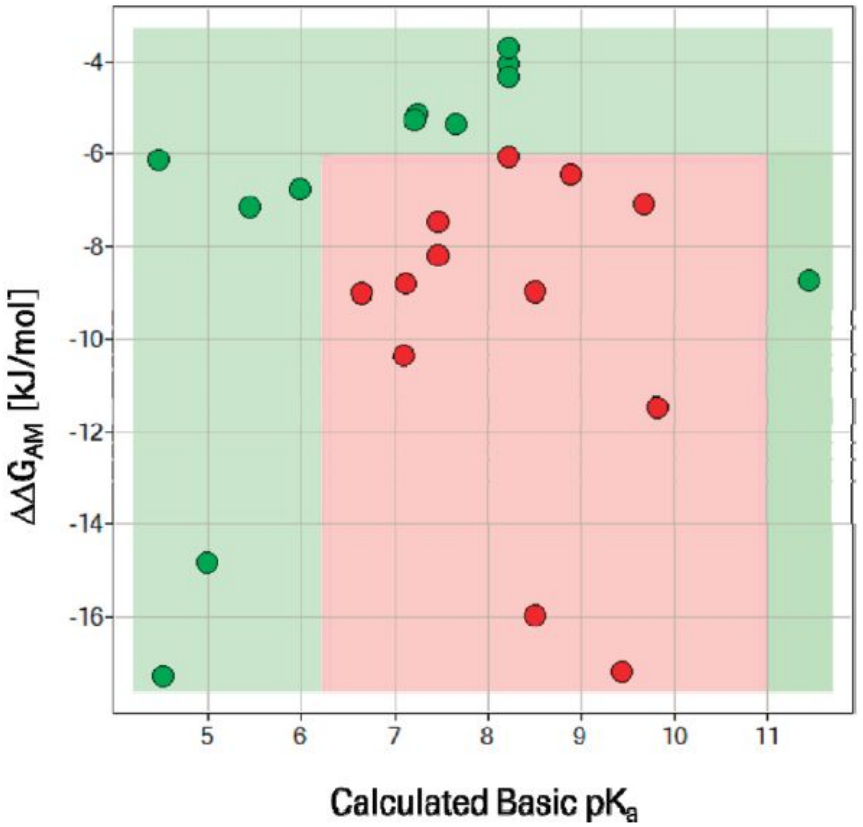




Iterative model building, experimentation, and model refining led to the predictive tool CAFCA

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

- 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!**