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 Lecture 5)

Q9: What statistical measure is used to measure the signal-noise ratio in screening? Can you use your own language explaining it? (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. On page 19: what is meant with "displacement of a labelled ligand"? (I do not know what 'displacement' means in that context)
2. I didn't quite understand the application of the Z value and when it usually is used
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

\[ L + R \xrightleftharpoons[k_2]{k_1} LR \]

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

At equilibrium, no net change of [LR]

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

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

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

\[ K_D = \frac{k_2}{k_1} \]

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

\[ [LR] = \frac{k_1[L][R_{total}]}{k_1[L] + k_2} \]
Four classical classes of mathematical models

**Compartment models**

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

Kinetics of ligand-target interaction

The Lotka-Volterra equations modelling predator-prey relationships.

\[ \frac{dx}{dt} = \alpha x - \beta xy, \]
\[ \frac{dy}{dt} = -\gamma y + \delta xy, \]
\[ \frac{dS}{dt} = -\frac{\beta IS}{N}, \]
\[ \frac{dI}{dt} = \frac{\beta IS}{N} - \gamma I, \]
\[ \frac{dR}{dt} = \gamma I \]

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**

Drug Delivery Device

Biological Tissue

Dissolution

Diffusion & Convection

Diffusion

Binding & Unbinding

**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 $K_D$

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

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

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

The law of mass action

\[
E + S \rightleftharpoons ES \rightarrow E + P
\]

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

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

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

Assuming that \( k_f[E][S] = k_r[ES] \)

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

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

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

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

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

\[
[ES] = \frac{k_f[E]_0[S]}{k_f(k_r/k_f + [S])}
\]
The dose-response curve and IC50: The Hill function and \textit{in vitro} pharmacology

- The Hill function is one of the mostly useful non-linear functions to model biological systems.
- In its general form, $H_{\text{max}}$ 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 $n=1$.

\[
H = H_{\text{max}} \frac{x^n}{k^n + x^n}
\]

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

Suppose it is an antiviral drug, compared with curve B, what does curve A, C, and D suggest?
More about the Hill function and dose-response curves

• The Hill function is often used to model either target occupancy or tissue response. In pharmacology, it is often used to model the tissue response.

• The Hill function can be approximated by a step function when $n$ goes towards infinity (top panel). This can be seen as one of the theoretical foundations of Boolean network modelling.

• The Hill function can be deduced from statistical mechanics of binding, a particle modelling approach. See for instance an article on Biophysics Wiki by Andreas Piehler for details.

• Data needs to be fit to the model, and in reality data can look quite different from the ideal curve (bottom panel). By setting priors, it is possible to perform inference even with ill-looking data.

The 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*.

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

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

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

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

- 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.

Interested in learning more about molecular modelling?

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


- A more advanced talk by Arthur Olson can be found [here](#), Workshop on the Mathematics of Drug Design/Discovery, June 4 - 8, 2018, The Fields Institute.

- Courses available at the University of Basel and beyond.
Molecular similarity and similarity measures

<table>
<thead>
<tr>
<th>Chemical similarity</th>
<th>A</th>
<th>B</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mol. weight</td>
<td>341.4</td>
<td>463.5</td>
</tr>
<tr>
<td>LogP</td>
<td>5.23</td>
<td>4.48</td>
</tr>
<tr>
<td>Rotatable bonds</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>Aromatic rings</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>Heavy atoms</td>
<td>26</td>
<td>35</td>
</tr>
</tbody>
</table>

Molecular similarity

2D similarity

3D similarity

Biological similarity

Global similarity

Local similarity

<table>
<thead>
<tr>
<th>Vascular endothelial growth factor receptor 2</th>
<th>Tyrosine-protein kinase TIE-2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Active</td>
<td>Inactive</td>
</tr>
</tbody>
</table>

| Manhattan distance | $D_{A,B} = \sum_{j=1}^{n} |x_{jA} - x_{jB}|$ | $D_{A,B} = a + b - 2c$ |
|--------------------|---------------------------|----------------------|
| Euclidean distance | $D_{A,B} = \left(\sum_{j=1}^{n} (x_{jA} - x_{jB})^2 \right)^{1/2}$ | $D_{A,B} = [a + b - 2c]^{1/2}$ |
| Cosine coefficient | $S_{A,B} = \frac{\sum_{j=1}^{n} x_{jA} x_{jB}}{\sqrt{\sum_{j=1}^{n} (x_{jA})^2 \sum_{j=1}^{n} (x_{jB})^2}}$ | \[ S_{A,B} = \frac{c}{\sqrt{a^2 + b^2}} \] |
| Dice coefficient   | $S_{A,B} = \frac{\sum_{j=1}^{n} x_{jA} x_{jB}}{\sum_{j=1}^{n} (x_{jA})^2 + \sum_{j=1}^{n} (x_{jB})^2}$ | $S_{A,B} = 2c/[a + b]$ |
| Tanimoto coefficient | $S_{A,B} = \frac{\sum_{j=1}^{n} x_{jA} x_{jB}}{\sum_{j=1}^{n} x_{jA} + \sum_{j=1}^{n} x_{jB}}$ | $S_{A,B} = c/[a + b - c]$ |
| Soergel distance$^b$ | $D_{A,B} = \left[ \sum_{j=1}^{n} \max(x_{jA}, x_{jB}) \right] / \left[ \sum_{j=1}^{n} \max(x_{jA}, x_{jB}) \right]$ | $D_{A,B} = 1 - \frac{c}{[a + b - c]}$ |

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

Molecular similarity does not equal biological similarity

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

Quantitative Structure-Activity Relationships (QSARs)

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

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.

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

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

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

Resources for learning about machine learning

**ESL** and **ISL**: From a frequentist view (almost)

**PRML** and **ITILA**: From a Bayesian view

**MLaPP**: Application oriented, more accessible, and balanced views

**Mathematical foundations**
Offline activities

• Read selected pages of *Computational Methods in Drug Discovery* by Sliwoski *et al*. Please submit your results to the Google Form, the link of which will be sent via a separate email.

• Optional and recommended:
  – Fill the anonymous survey #5 (link will be sent via a separate email).
  – Recommended readings:
More about the the Free-Wilson analysis


- A Python implementation on GitHub, and a blog post going through examples, is shared by Pat Walters.

- Free-Wilson nonadditivity is a research topic, for instance see Cramer et al., 2015

- Source of the example shown in the lecture: QSAR of the ACCVIP project (The Australian Computational Chemistry via the Internet Project)
Drug-induced phospholipidosis is correlated with amphiphilicity

- Phospholipidosis is a lysosomal storage disorder characterized by the excess accumulation of phospholipids in tissues.

- Drug-induced phospholipidosis is caused by cationic amphiphilic drugs and some cationic hydrophilic drugs.

- Clinical pharmacokinetic characteristics of drug-induced phospholipidosis include (1) very long terminal half lives, (2) high volume of distribution, (3) tissue accumulation upon frequent dosing, and (4) deficit in drug metabolism.

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

\[ \hat{A} = \sum_{i} d \cdot \hat{\alpha}_i \]

\( \hat{A} \): Calculated amphiphilic moment
\( d \): distance between the center of gravity of the charged part of a molecule and the hydrophobic/hydrophilic remnant of the molecule
\( \hat{\alpha}_i \): 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.

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

<table>
<thead>
<tr>
<th>in vitro/in vivo</th>
<th>in silico/in vivo</th>
<th>Exp. PC/in vivo</th>
<th>In silico/in vitro</th>
<th>n=36</th>
</tr>
</thead>
<tbody>
<tr>
<td>94%</td>
<td>81%</td>
<td>89%</td>
<td>89%</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>in vitro/in silico</th>
<th>n=422</th>
</tr>
</thead>
<tbody>
<tr>
<td>Accuracy [(TP+TN)/(P+N)]</td>
<td>Sensitivity [True Positive Rate]</td>
</tr>
<tr>
<td>86%</td>
<td>80%</td>
</tr>
</tbody>
</table>

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!