The study of biomolecular binding equilibria

THE FOCUS OF THIS TRAINING DAY IS DATA ANALYSIS

- choose experimental method
- choose concentrations
- collect data
- data analysis

- preliminary vs. refinement experiments
- publish report
- binding constants
- molecular mechanism

Part 1: Theory

Petr Kuzmič
BioKin, Ltd.
# Numerical vs. algebraic fitting models

<table>
<thead>
<tr>
<th>Algebraic fitting models</th>
<th>Numerical fitting models</th>
</tr>
</thead>
<tbody>
<tr>
<td>single algebraic equations</td>
<td>systems of simultaneous equations</td>
</tr>
<tr>
<td>may not exist for some mechanisms</td>
<td>always exist for any mechanism</td>
</tr>
<tr>
<td>must be derived by hand</td>
<td>derived automatically by the computer</td>
</tr>
<tr>
<td>special experimental conditions</td>
<td>applicable to any set of conditions</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>many software packages</th>
<th>SigmaPlot, GraphPad, Origin, ...</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>highly specialized software DynaFit, BioEQS</td>
</tr>
</tbody>
</table>

## ADVANTAGES OF THE GENERAL NUMERICAL APPROACH

### The DynaFit software package

**ONLY A QUICK GLANCE DURING THIS IS A TRAINING DAY, NOT A THEORY CLASS**

### REFERENCES


### CITATION ANALYSIS

- Cited in approximately **850** journal articles since 1998
- Journals most frequently citing DynaFit: *Biochemistry, J. Biol. Chem.*

### WHAT CAN DYNAFIT DO FOR YOU

- Derive mathematical models for data fitting, fully automatically.

---
Example 1: Competitive ligand displacement assay

This problem can be handled algebraically, although it is a stretch...

An exact mathematical expression for describing competitive binding of two different ligands to a protein molecule

Zhi-Xin Wang*
National Laboratory of Biomacromolecules, Institute of Biophysics, Academia Sinica, Beijing 100101, China
Received 30 December 1994

Algebraic data-fitting model:

\[
[P] = \frac{a}{3} + \frac{2}{3} \sqrt{a^2 - 3b} \cos \theta
\]

\[
[PA] = \frac{[A]_0 [2(\sqrt{a^2 - 3b}) \cos(\theta/3) - a]}{3K_A + [2(\sqrt{a^2 - 3b}) \cos(\theta/3) - a]}
\]

\[
[PB] = \frac{[B]_0 [2(\sqrt{a^2 - 3b}) \cos(\theta/3) - a]}{3K_B + [2(\sqrt{a^2 - 3b}) \cos(\theta/3) - a]}
\]

where

\[
\theta = \arccos \left( \frac{-2a^3 + 9ab - 27c}{2\sqrt{(a^2 - 3b)^3}} \right)
\]

\[
a = K_A + K_B + [A]_0 + [B]_0 - [P]_0
\]

\[
b = K_A([A]_0 - [P]_0) + K_B([B]_0 - [P]_0) + K_A K_B
\]

\[
c = -K_A K_B [P]_0
\]

Competitive ligand displacement in DynaFit

This problem can be handled algebraically, although it is a stretch...

An exact mathematical expression for describing competitive binding of two different ligands to a protein molecule

Zhi-Xin Wang*
National Laboratory of Biomacromolecules, Institute of Biophysics, Academia Sinica, Beijing 100101, China
Received 30 December 1994

DynaFit data-fitting model:

\[
[P] = \frac{a}{3} + \frac{2}{3} \sqrt{a^2 - 3b} \cos \theta
\]

\[
[PA] = \frac{[A]_0 [2(\sqrt{a^2 - 3b}) \cos(\theta/3) - a]}{3K_A + [2(\sqrt{a^2 - 3b}) \cos(\theta/3) - a]}
\]

\[
[PB] = \frac{[B]_0 [2(\sqrt{a^2 - 3b}) \cos(\theta/3) - a]}{3K_B + [2(\sqrt{a^2 - 3b}) \cos(\theta/3) - a]}
\]

where the requisite mathematics is “somehow” handled by the computer
Example 2: A “complex” binding mechanism

This problem cannot be handled algebraically, even in principle!

Stoichiometry and affinity for thymine DNA glycosylase binding to specific and nonspecific DNA

Michael T. Morgan, Atanu Maiti, Megan E. Fitzgerald and Alexander C. Drohat

There can be no algebraic fitting model for this experiment!

A “complex” binding mechanism in DynaFit

This problem cannot be handled algebraically, even in principle!

DynaFit data-fitting model:

[mechanism]

\[
\begin{align*}
\text{DNA} + \text{Gly} & \rightleftharpoons \text{DNA.Gly} \\
\text{DNA.Gly} + \text{Gly} & \rightleftharpoons \text{Gly.DNA.Gly} \\
\text{DNA} + \text{Comp} & \rightleftharpoons \text{DNA.Comp} \\
\text{DNA.Comp} + \text{Comp} & \rightleftharpoons \text{Comp.DNA.Comp}
\end{align*}
\]

\[
\begin{align*}
\text{Kd1} & \quad \text{dissoc} \\
\text{Kd2} & \quad \text{dissoc} \\
\text{Kd1}^* & \quad \text{dissoc} \\
\text{Kd2}^* & \quad \text{dissoc}
\end{align*}
\]
Theoretical considerations

**MINIMUM AMOUNT OF THEORY NEEDED FOR CONSTRUCTING MECHANISMS IN DYNAFIT**

- Statistical factors
- Thermodynamic box
- Intensive physical quantities
- Rapid equilibrium enzyme kinetics

Single-site and multi-site binding

"P" = PROTEIN, "L" = LIGAND. OTHER SYMBOLS WOULD WORK EQUALLY WELL

**Single-site binding:** one complex formed

\[
[P + L \leftrightarrow [P:L] : K_d \text{ dissoc}]
\]

**Two-site binding:** two complexes formed

\[
\begin{align*}
[P + L \leftrightarrow [P:L] : K_{d1} \text{ dissoc} \\
[P:L + L \leftrightarrow [P:L2] : K_{d2} \text{ dissoc}
\end{align*}
\]
**Cooperativity in multi-site binding**

VALUES "12.34" AND "56.78" STAND FOR ANY SUITABLY CHOSEN NUMERICAL VALUE

Two **non-interacting** sites: *one adjustable* $K_d$ value

**[mechanism]**

\[
P + L \iff P.L : K_d \text{ dissoc}
\]

\[
P.L + L \iff P.L2 : K_d \text{ dissoc}
\]

**[constants]**

- $K_d = 12.34$ \text{ ?} ; optimized parameter
- $K_d = 4 \times K_d$ \text{ statistical factor}

Two **cooperative** sites: *two adjustable* $K_d$ values

**[mechanism]**

\[
P + L \iff P.L : K_d \text{ dissoc}
\]

\[
P.L + L \iff P.L2 : K_d \text{ dissoc}
\]

**[constants]**

- $K_d = 12.34$ \text{ ?} ; optimized parameter
- $K_d = 56.78$ \text{ ?} ; optimized parameter

---

**Association step: Statistical factors**

FOR IDENTICAL NON-INTERACTING SITES, $P \to \text{P.L}$ IS **TWICE AS LIKELY TO OCCUR** AS $\text{P.L} \to \text{L.P.L}$

\[k_{on}^{(1)} \quad k_{on}^{(2)}\]

**two possibilities** for the first "marble" to fall in

\[k_{on}^{(1)} \quad k_{on}^{(2)}\]

**one possibility** for the second "marble" to fall in

Rate constants: \[k_{on}^{(1)} = 2 \times k_{on}^{(2)}\]
**Dissociation step: Statistical factors**

For identical non-interacting sites, L.P.L $\rightarrow$ P.L is **twice as likely to occur** as P.L $\rightarrow$ P.

- **One possibility** for the remaining “marble” to fall out
- **Two possibilities** for a “marble” to fall out

**Rate constants:**

$$k_{(1)}^{(2)} = 2 \times k_{(1)}^{(1)}$$

---

**Equilibrium: Statistical factors**

For two identical non-interacting sites, $K_d^{(2)}$ is **four times larger** than $K_d^{(1)}$.

- **Recall:**
  - $k_{on}^{(2)} = 2 \times k_{on}^{(1)}$
  - $k_{off}^{(2)} = k_{off}^{(1)} / 2$

**Kd (1) and Kd (2) relations:**

$$K_d^{(1)} = \frac{k_{off}^{(1)}}{k_{on}^{(1)}}$$

$$K_d^{(2)} = \frac{k_{off}^{(2)}}{k_{on}^{(2)}}$$

**Expression for Kd (2):**

$$K_d^{(2)} = \frac{k_{off}^{(1)} \times 2}{k_{on}^{(1)} / 2} = \frac{K_d^{(1)} \times 2}{1 / 2} = K_d^{(1)} \times 4$$
Statistical factors for two binding sites - Summary

For two identical, non-interacting sites, $K_d^{(2)}$ is always four times larger than $K_d^{(1)}$.

For two cooperative sites, both $K_d^{(2)}$ and $K_d^{(1)}$ can attain any arbitrary values.

DYNAFIT NOTATION FOR NON-INTERACTING SITES

[mechanism]
\[
P + L \leftrightarrow P \cdot L : \quad K_d \text{ dissoc} \\
P \cdot L + L \leftrightarrow P \cdot L_2 : \quad K_d \text{ dissoc}
\]

[constants]
\[
K_d^{(1)} = \ldots ? ; \text{ any appropriate value} \\
K_d^{(2)} = 4 \times K_d^{(1)} ; \text{ statistical factor}
\]

Example:
\[
\begin{align*}
K_d^{(1)} &= \frac{K_d}{n} \\
K_d^{(2)} &= \frac{2}{3} K_d \\
K_d^{(3)} &= \frac{3}{2} K_d \\
K_d^{(4)} &= K_d
\end{align*}
\]

$K_d^{(1)} = \frac{4}{1} K_d$
$K_d^{(2)} = \frac{2}{1} K_d$
$K_d^{(3)} = \frac{3}{2} K_d$
$K_d^{(4)} = K_d$

\[
1 : 2.66667 : 6 : 16
\]

Statistical factors for multiple identical binding sites


\[
K_d^{(i)} = K_d \frac{i}{n - i + 1}
\]

$n$ = number of binding sites
\[
i \quad = \text{ith binding step}
\]
\[
K_d = \text{microscopic dissociation constant}
\]
\[
K_d^{(i)} = \text{macroscopic } K_d \text{ in } i\text{th binding step}
\]

Example: $n = 4$
\[
\begin{align*}
K_d^{(1)} &= \frac{1}{4} K_d \\
K_d^{(2)} &= \frac{2}{3} K_d \\
K_d^{(3)} &= \frac{3}{2} K_d \\
K_d^{(4)} &= K_d
\end{align*}
\]

\[
1 : 2.66667 : 6 : 16
\]
Statistical factors in DynaFit

distributed example file: ./courses/BSTD-2014/ThT_22AG/noninteracting.txt

Gabelica et al. (2013) Biochemistry 2013, 52, 5620-5628, Figure 3D

ASSUME FOUR IDENTICAL, NON-INTERACTING DNA/Ligand SITES

[mechanism]

DNA       + Lig <==> DNA.Lig :     Kd1    dissoc
DNA.Lig   + Lig <==> DNA.Lig.2 :     Kd2    dissoc
DNA.Lig.2 + Lig <==> DNA.Lig.3 :     Kd3    dissoc
DNA.Lig.3 + Lig <==> DNA.Lig.4 :     Kd4    dissoc

[constants]

Kd1 =       40 ?       ; = 1/4 Kd , Kd ... microscopic
Kd2 = 2.66667 * Kd1   ; = 2/3 Kd
Kd3 = 6 * Kd1         ; = 3/2 Kd
Kd4 = 16 * Kd1        ; = 4   Kd

Theoretical considerations

MINIMUM AMOUNT OF THEORY NEEDED FOR CONSTRUCTING MECHANISMS IN DYNAFIT

- Statistical factors
- Thermodynamic box
- Intensive physical quantities
- Rapid equilibrium enzyme kinetics
**Cycles in binding mechanisms: Thermodynamic box**

**Example:**
enzyme "E" simultaneously binding two co-substrates "A" and "B"

\[
\begin{align*}
E & \quad \text{thermodynamic box} \quad E \cdot A \quad E \cdot B \quad \rightarrow \quad E \cdot A \cdot B \quad \rightarrow \quad \text{products}
\end{align*}
\]

**Conservation of energy: Overall \( K_{eq} \) must be unity**

clockwise around the cycle:

\[
\frac{1}{K_d^{(A)}} \times \frac{1}{K_d^{(AB)}} \times K_d^{(B)} \times K_d^{(BA)} = 1
\]

upper branch must meet lower branch:

\[
K_d^{(A)} \times K_d^{(AB)} = K_d^{(B)} \times K_d^{(BA)}
\]
The “leave one out” rule for thermodynamic boxes

**How to represent cyclic binding mechanisms in Dynafit**

There are four equivalent ways to represent this mechanism.

Number of binding steps must match the number of unique complexes.

Always check published literature results

Competitive and Cooperative Interactions Mediate RNA Transfer from Herpesvirus Saimiri ORF57 to the Mammalian Export Adaptor ALYREF

Richard B. Tannicifliffa, Guillaume M. Hautbergueb, Stuart A. Wilsonb, Priti Kakac, Alexander P. Golovanov1,2

left branch must meet right branch:

\[ K_d^{(A)} \times K_d^{(AB)} = K_d^{(B)} \times K_d^{(BA)} \]
Theoretical considerations

MINIMUM AMOUNT OF THEORY NEEDED FOR CONSTRUCTING MECHANISMS IN DYNAFIT

- Statistical factors
- Thermodynamic box
- Intensive physical quantities
- Rapid equilibrium enzyme kinetics

Two types of observable physical quantities

<table>
<thead>
<tr>
<th>Extensive</th>
<th>Intensive</th>
</tr>
</thead>
<tbody>
<tr>
<td>Signal is proportional to concentrations</td>
<td>Signal is proportional to mole fractions</td>
</tr>
<tr>
<td>fluorescence intensity</td>
<td>fluorescence polarization (anisotropy)</td>
</tr>
<tr>
<td>NMR peak area</td>
<td>NMR chemical shift</td>
</tr>
<tr>
<td>UV/Vis absorbance</td>
<td>...</td>
</tr>
<tr>
<td>HPLC peak area</td>
<td></td>
</tr>
<tr>
<td>radioactive counts</td>
<td></td>
</tr>
<tr>
<td>optical rotation</td>
<td></td>
</tr>
<tr>
<td>...</td>
<td></td>
</tr>
</tbody>
</table>

Binding Constants & Mechanisms pt.1
HOW TO REPRESENT INTENSIVE VARIABLES IN DYNAFIT

Use the DynaFit keyword "intensive" in the [responses] section of the input script:

```plaintext
[responses]
intensive
...
```

**Example:** Protein-protein binding constants determined by NMR

![Graph]

Figure 10.1 NMR chemical shift titration of the PRDM5 protein (total concentration varied between 0.125 and 0.1172 mM) with a model peptide ligand.


---

**Theoretical considerations**

**MINIMUM AMOUNT OF THEORY NEEDED FOR CONSTRUCTING MECHANISMS IN DYNAFIT**

- Statistical factors
- Thermodynamic box
- Intensive physical quantities
- Rapid equilibrium enzyme kinetics
Rapid-equilibrium approximation in enzyme kinetics


The Michaelis-Menten mechanism and rate equation:

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

\[
\frac{\text{rate}}{v} = \frac{[S]}{K_S + [S]}
\]

How is this derived?

The rate is equal to the concentrations of all product-forming species, each multiplied by its catalytic rate constant. \( v = k_p [ES] \)

Rate is proportional to the equilibrium concentrations of reactive complexes!

Enzyme kinetics treated as simple “binding equilibria”

1. Compute the composition at equilibrium.
2. Look up all enzyme-substrate complexes that do form products.
3. Multiply their concentrations by an appropriate proportionality constant:
   \[
   \text{constant} = \text{molar instrumental response of the product} \times \text{relevant } k_{\text{cat}}
   \]
4. Compute the sum total of all such terms.

The result is the initial rate under the rapid equilibrium approximation.
Rapid-equilibrium enzyme kinetics in DynaFit

TWO EQUIVALENT WAYS TO REPRESENT RAPID-EQUILIBRIUM ENZYME KINETICS

DYNAFIT

See "DynaFit Scripting Manual" on http://www.biokin.com/

**METHOD 1**: initial rate formalism

```plaintext
{task}
data = rates
    approximation = rapid-equilibrium

{mechanism}
E + S <=> E.S : Ks  dissoc
    E.S ----> E + P : kcat

{constants}
    Ks = ...
    kcat = 3

{responses}
    P = 4
    ...
```

**METHOD 2**: equilibrium formalism

```plaintext
{task}
data = equilibria

{mechanism}
E + S <=> E.S : Ks  dissoc
    E.S ----> E + P : kcat

{constants}
    Ks = ...

{responses}
    E.S = 12  ;  = 3 x 4
    ...
```

**Summary**

- **Statistical factors**
  - Independent binding sites: \( K_s \)s are linked via statistical factors.
  - Cooperative binding sites: \( K_s \)s can attain arbitrary values.

- **Thermodynamic boxes**
  - The "leave one out" rule: thermodynamic cycles must remain open. It does not matter which edge of the box is left out.

- **Intensive physical quantities**
  - Use *intensive* keyword for NMR chemical shift or fluorescence polarization. Omit this keyword for fluorescence *intensity*, UV/Vis absorbance, etc.

- **Rapid equilibrium enzyme kinetics**
  - All rapid equilibrium enzyme kinetics can be expressed as “binding equilibria”. Turnover numbers ("kcat" values) become “responses” in the binding model.