Determination of Binding Affinities
and Molecular Mechanisms

Biochemical Society
Advancing Molecular Bioscience

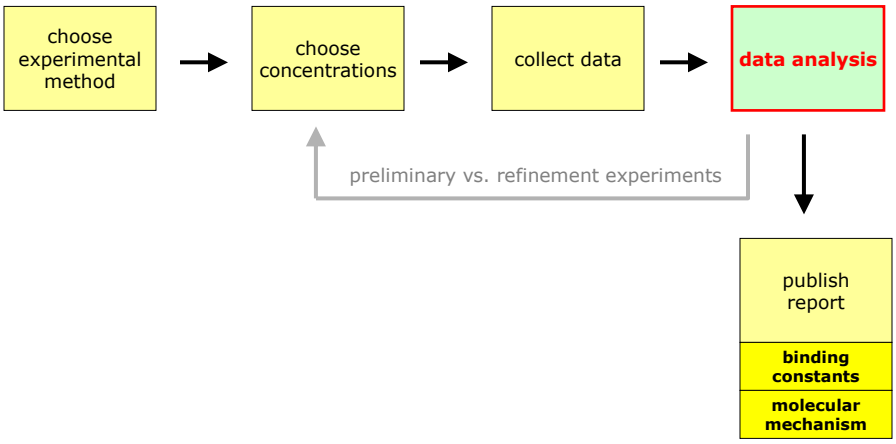
Training Day
May 2, 2014 (London)

Part 1: Theory

Petr Kuzmič
BioKin, Ltd.

The study of biomolecular binding equilibria

THE FOCUS OF THIS TRAINING DAY IS **DATA ANALYSIS**



Numerical vs. algebraic fitting models

ADVANTAGES OF THE GENERAL **NUMERICAL** APPROACH

Algebraic fitting models	**Numerical** fitting models
single algebraic equations	systems of simultaneous equations
may not exist for some mechanisms	always exist for any mechanism
must be derived by hand	derived automatically by the computer
special experimental conditions	applicable to any set of conditions
many software packages ***SigmaPlot, GraphPad, Origin,*** *...*	highly specialized software ***DynaFit****,* ***BioEQS***

The DynaFit software package

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

REFERENCES

1. Kuzmic, P. (1996) *Anal. Biochem.* **237**, 260-273
2. Kuzmic, P. (2009) *Meth. Enzymol.* **467**, 247-280

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...

FEBS 15134 FEBS Letters 360 (1995) 111–114

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

$$P + A \underset{}{\overset{K_A}{\rightleftharpoons}} PA, \qquad P + B \overset{K_B}{\rightleftharpoons} PB$$

Algebraic data-fitting model:

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

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

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

where

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

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

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

$$c = -K_AK_B[P]_0$$

Competitive ligand displacement in DynaFit

THIS PROBLEM **CAN** BE HANDLED ALGEBRAICALLY, ALTHOUGH IT IS A STRETCH...

FEBS 15134 FEBS Letters 360 (1995) 111–114

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

$P + A \overset{K_A}{\rightleftharpoons} PA, \qquad P + B \overset{K_B}{\rightleftharpoons} PB$

DynaFit data-fitting model:

```
[mechanism]
   P + A <==> PA   :    Ka   dissociation
   P + B <==> PB   :    Kb   dissociation
```

where the requisite mathematics is "somehow" handled by the computer

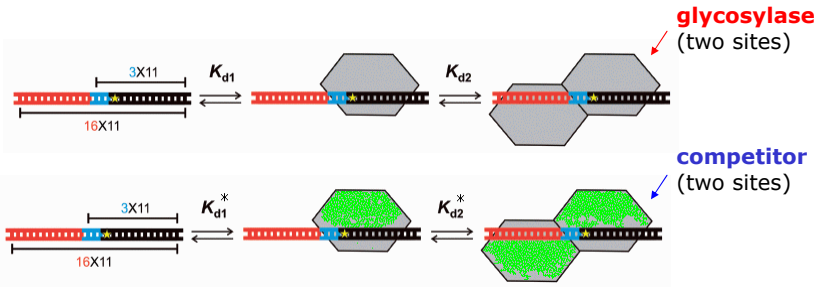
Example 2: A “complex” binding mechanism

THIS PROBLEM **CANNOT** BE HANDLED ALGEBRAICALLY, EVEN IN PRINCIPLE!

Published online 21 November 2010 *Nucleic Acids Research, 2011, Vol. 39, No. 6* **2319–2329**
doi:10.1093/nar/gkq1164

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

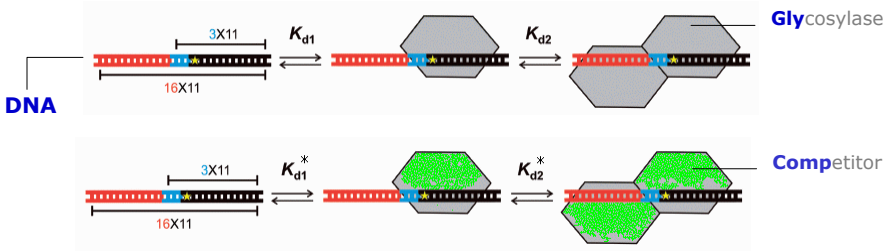
Michael T. Morgan, Atanu Maiti, Megan E. Fitzgerald and Alexander C. Drohat*



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]
   DNA + Gly <==> DNA.Gly                :    Kd1   dissoc
   DNA.Gly + Gly <==> Gly.DNA.Gly        :    Kd2   dissoc

   DNA + Comp <==> DNA.Comp              :    Kd1*  dissoc
   DNA.Comp + Comp <==> Comp.DNA.Comp    :    Kd2*  dissoc
```

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*

```
[mechanism]
   P + L <==> P.L    :  Kd   dissoc
```

Two-site binding: *two complexes formed*

```
[mechanism]
   P + L <==> P.L       :    Kd1   dissoc
   P.L + L <==> P.L2    :    Kd2   dissoc
```

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 <==> P.L       :    Kd1   dissoc
   P.L + L <==> P.L2    :    Kd2   dissoc

[constants]
   Kd1 = 12.34 ?     ; optimized parameter
   Kd2 = 4 * Kd1     ; statistical factor
```

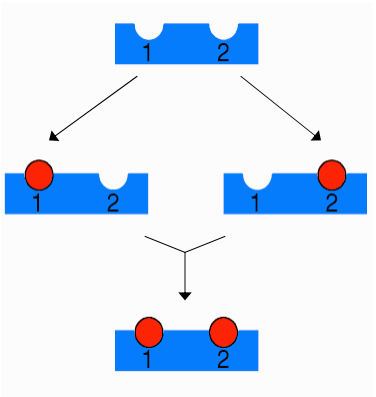
Two **cooperative** sites: *two adjustable K_d values*

```
[mechanism]
   P + L <==> P.L       :    Kd1   dissoc
   P.L + L <==> P.L2    :    Kd2   dissoc

[constants]
   Kd1 = 12.34 ?     ; optimized parameter
   Kd2 = 56.78 ?     ; optimized parameter
```


Association step: Statistical factors

FOR IDENTICAL NON-INTERACTING SITES, P → P.L IS **TWICE AS LIKELY TO OCCUR** AS P.L → L.P.L



$k_{on}^{(1)}$

two possibilities
for the first "marble" to fall in

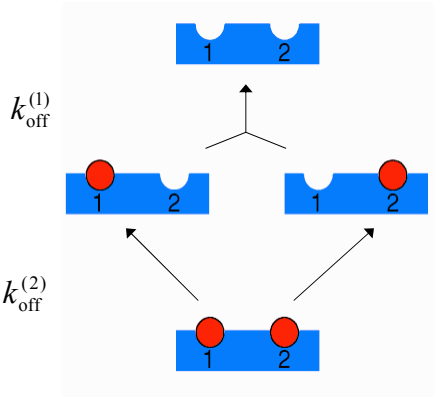
$k_{on}^{(2)}$

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 → P.L IS **TWICE AS LIKELY TO OCCUR** AS P.L → P



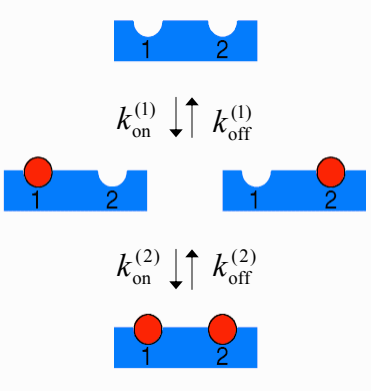
one possibility
for the remaining "marble" to fall out

two possibilities
for *a* "marble" to fall out

Rate constants: $k_{off}^{(2)} = 2 \times k_{off}^{(1)}$

Equilibrium: Statistical factors

FOR TWO IDENTICAL NON-INTERACTING SITES, $K_d^{(2)}$ IS **FOUR TIMES LARGER** THAN $K_d^{(1)}$



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

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

recall:

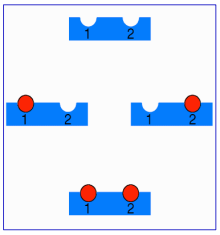
$$k_{on}^{(1)} = 2 \times k_{on}^{(2)}$$

$$k_{on}^{(2)} = k_{on}^{(1)} / 2$$

$$K_d^{(2)} = \frac{k_{off}^{(1)} \times 2}{k_{on}^{(1)} / 2} = K_d^{(1)} \frac{2}{1/2} = K_d^{(1)} \times 4$$

Statistical factors for two binding sites - Summary

HOW TO REPRESENT (LACK OF) COOPERATIVITY IN **DYNAFIT**



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 <==> P.L       :    Kd1   dissoc
   P.L + L <==> P.L2    :    Kd2   dissoc

[constants]
   Kd1 = ... ?       ; any appropriate value
   Kd2 = 4 * Kd1     ; statistical factor
```


Statistical factors for multiple identical binding sites

Bisswanger, H. (2008) *Enzyme Kinetics*, 2nd Ed., Wiley-VCH, Tuebingen, p. 14, Eq. (2)

NON-INTERACTING, IDENTICAL BINDING SITES

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

n	= number of binding sites
i	= ith binding step
K_d	= microscopic dissociation constant
$K_d^{(i)}$	= macroscopic K_d in ith binding step

EXAMPLE: ***n* = 4**

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

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

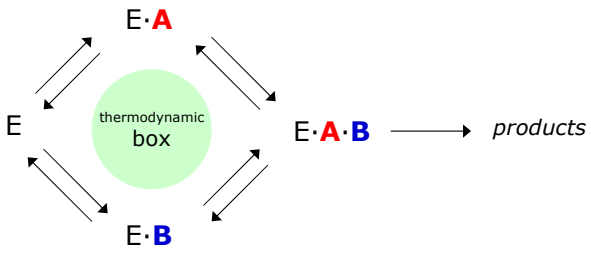
MINIMUM AMOUNT OF THEORY NEEDED FOR CONSTRUCTING MECHANISMS IN DYNAFIT

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

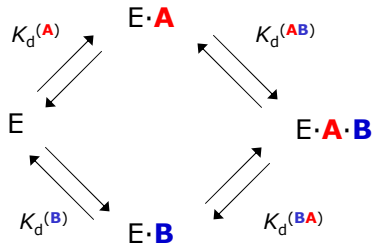
Cycles in binding mechanisms: Thermodynamic box

Example:

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



Conservation of energy: Overall K_{eq} must be unity

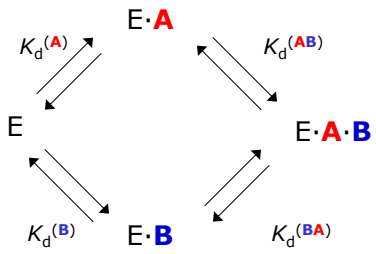


clockwise around the cycle: $$\frac{1}{K_d^{(A)}} \times \frac{1}{K_d^{(AB)}} \times K_d^{(B)} \times K_d^{(BA)} = 1$$

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**




```
[mechanism]
   E + A <==> E.A      : KdA   dissoc
   E + B <==> E.B      : KdB   dissoc
   E.A + B <==> E.A.B  : KdAB  dissoc
```

```
[mechanism]
   E + A <==> E.A      : KdA   dissoc
   E + B <==> E.B      : KdB   dissoc
   E.B + A <==> E.A.B  : KdBA  dissoc
```

```
[mechanism]
   E + A <==> E.A      : KdA   dissoc
   E.A + B <==> E.A.B  : KdAB  dissoc
   E.A.B <==> E.B + A  : KdBA  dissoc
```

```
[mechanism]
   E + B <==> E.B      : KdB   dissoc
   E.B + A <==> E.A.B  : KdBA   dissoc
   E.A.B <==> E.A + B  : KdAB  dissoc
```

There are **four equivalent ways** to represent this mechanism.

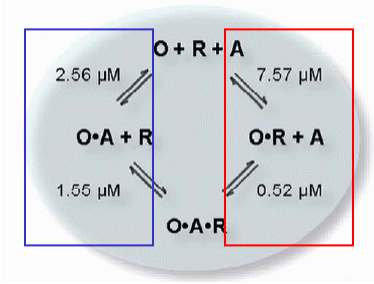
Number of binding steps must match the number of unique complexes.

Always check published literature results

OPEN ACCESS Freely available online PLoS Pathog 10(2): e1003907 (2014) PLOS | PATHOGENS

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

Richard B. Tunnicliffe[1], Guillaume M. Hautbergue[2], Stuart A. Wilson[3], Priti Kalra[1], Alexander P. Golovanov[1]*



2.56 × 1.55 = **3.97**

7.57 × 0.52 = **3.93** ✓

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

Extensive	**Intensive**
Signal is proportional to concentrations	Signal is proportional to mole fractions
fluorescence intensity NMR peak area UV/Vis absorbance HPLC peak area radioactive counts optical rotation ...	fluorescence polarization (anisotropy) NMR chemical shift ...

Intensive physical variables in DynaFit

HOW TO REPRESENT INTENSIVE VARIABLES IN **DYNAFIT**

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

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

Example: Protein-protein binding constants determined by NMR

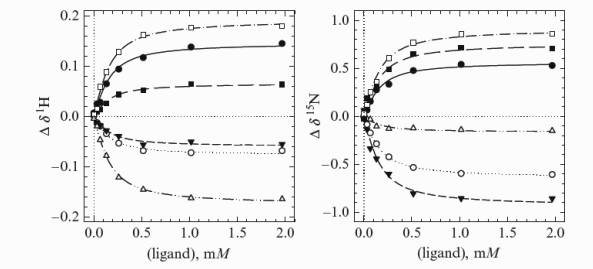


Figure 10.1 NMR chemical shift titration of the PRDM5 protein (total concentration varied between 0.125 and 0.1172 m*M*) with a model peptide ligand.

Kuzmic, P. (2009) *Meth. Enzymol.* **467**, 247-280

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

I. Segel (1975) "Enzyme Kinetics", J. Wiley, New York, pp. 22-24

The Michaelis-Menten mechanism and **rate** equation:

$$E + S \underset{}{\overset{K_S}{\rightleftharpoons}} ES \xrightarrow{k_p} E + P$$

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

How is this derived?

v 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:

   constant = molar instrumental response of the product × relevant k_{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

```
[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

```
[task]
   data = equilibria


[mechanism]
   E + S <==> E.S  :  Ks    dissoc


[constants]
   Ks = ...


[responses]
   E.S = 12     ;     = 3 x 4

...
```


Summary

- Statistical factors

  Independent binding sites: K_ds are linked via statistical factors.
  Cooperative binding sites: K_ds 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 ("k_{cat}" values) become "responses" in the binding model.