MASARYKOVA UNIVERZITA
INNOVATION LECTURES (INNOIEC)
www.muni.cz

Binding and Kinetics for Experimental Biologists

Lecture 1

Numerical Models for Biomolecular Interactions

Petr Kuzmič, Ph.D.
BioKin, Ltd.

WATERTOWN, MASSACHUSETTS, U.S.A.

Tento projekt je spolufinancován Evropským sociálním fondem a státním rozpočtem České republiky.

EVROPSKÁ UNIE esf
MINISTERSTVO ŠKOLSTVÍ, MLÁDEŽE A TĚLOVÝCHOVY
OP Vzdělávání pro konkurenceschopnost
UNIVERSITAS MASARYKIANA BRUNENSIS
INVESTICE DO ROZVOJE VZDĚLÁVÁNÍ

BioKin L t d.

Lecture outline

- The problem:

  Traditional equations for fitting biomolecular binding data restrict the experimental design. Typically, at least one component must be present in very large excess.

- The solution:

  Abandon algebraic equations entirely. Use iterative numerical models, which can be derived automatically by the computer.

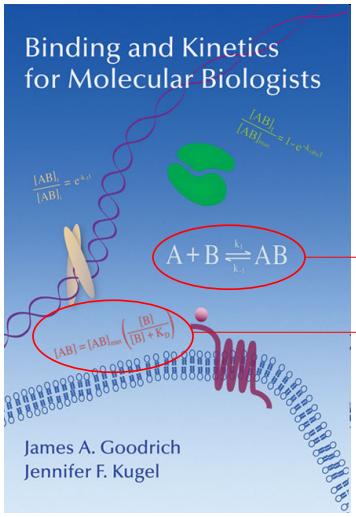
- An implementation:

  Software *DynaFit*.

- An example:

  Kinetics of forked DNA binding to the protein-protein complex formed by DNA-polymerase sliding clamp (gp45) and clamp loader (gp44/62).

Traditional approach is based on algebraic models



A typical "cookbook"
for experimental biologists

molecular mechanism

algebraic model

Cold Spring Harbor Laboratory Press
Cold Spring Harbor, NY, 2007
ISBN-10: 0879697369

Algebraic models restrict experiment design

Goodrich & Kugel (2006) ***Binding and Kinetics for Molecular Biologists***

EXAMPLE: Determine the association rate constant for **A + B → AB**

Summary: How to experimentally measure k_1

1. Develop an assay to monitor AB.
2. Combine A and B at $t = 0$ ($[B]_i >> [AB]_{max}$).
3. Measure accumulation of AB over time.
4. Plot data and determine k_{obs}
5. Repeat Steps 2–4 at multiple different $[B]_i$.
6. Plot k_{obs} versus $[B]_i$ and determine k_1
7. Consider whether k_1 is diffusion limited.

$$AB_t = AB_{max}(1 - e^{-k_{obs}t})$$

$$k_{obs} = k_1[B]_i + k_{-1}$$

Typically, this condition is met by setting the initial concentration of B to be much greater than the initial concentration of A (e.g., 100-fold).

Experimental handbooks are full of restrictions

Goodrich & Kugel (2007) ***Binding and Kinetics for Molecular Biologists***

> Under conditions where the concentration of B is much greater than $K_{D(AB)}$ ([B] >> $K_{D(AB)}$) the following equation can be used

p. 34

> In using this technique, it is important to make sure that the concentration of unlabeled A is greater than the concentration of B in the reactions.

p. 79

> Third, reactions should be performed under conditions where the initial concentration of B ($[B]_i$) is much greater than the sum of the final concentrations of AB and AB*

p. 120

> Using the abortive initiation assay, k_{obs} values were measured at multiple different RNA polymerase concentrations (always greater than the concentration of DNA) at two different promoters.

p. 126

etc.

Numerical vs. algebraic mathematical models

FROM A VARIETY OF ALGEBRAIC EQUATIONS TO A UNIFORM SYSTEM OF **DIFFERENTIAL EQUATIONS**

EXAMPLE: Determine the rate constant k_1 and k_{-1} for **A + B** $\underset{k_{-1}}{\overset{k_1}{\rightleftharpoons}}$ **AB**

ALGEBRAIC EQUATIONS	DIFFERENTIAL EQUATIONS
$AB_i = AB_{max}(1 - e^{-k_{obs}t})$ $k_{obs} = k_1[B]_i + k_{-1}$	$d[A]/dt = -k_1[A][B] + k_{-1}[AB]$ $d[B]/dt = -k_1[A][B] + k_{-1}[AB]$ $d[AB]/dt = +k_1[A][B] - k_{-1}[AB]$
Applies only when [B] >> [A]	Applies under all conditions

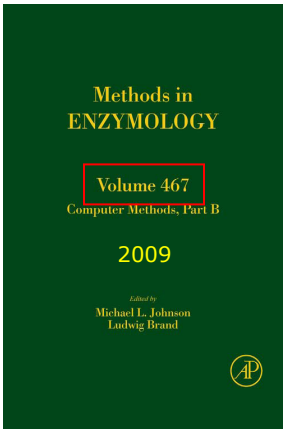
Advantages and disadvantages of numerical models

THERE IS NO SUCH THING AS A FREE LUNCH

ADVANTAGE	ALGEBRAIC MODEL	DIFFERENTIAL MODEL
can be derived for any molecular mechanism	-	+
can be derived automatically by computer	-	+
can be applied under any experimental conditions	-	+
can be evaluated without specialized software	+	-
requires very little computation time	+	-
does not require an initial estimate	+	-
is resistant to truncation and round-off errors	+	-
has a long tradition: many papers published	+	-

Specialized numerical software: *DynaFit*

MORE THAN **600 PAPERS** PUBLISHED WITH IT (1996 – 2009)



CHAPTER TEN

DYNAFIT—A SOFTWARE PACKAGE FOR ENZYMOLOGY

Petr Kuzmič

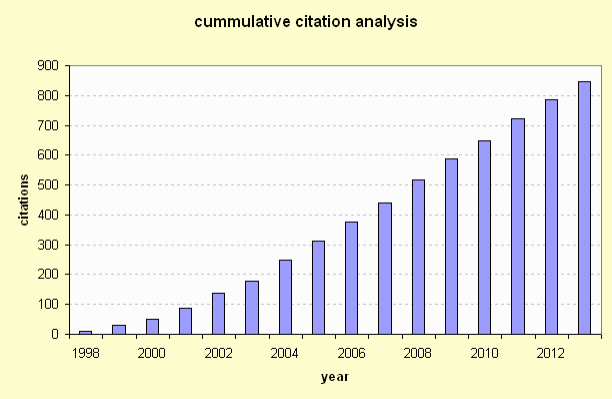
DOWNLOAD http://www.biokin.com/dynafit

FREE for academic users

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

DynaFit software: Citation analysis

APPROXIMATELY 850 JOURNAL ARTICLES PUBLISHED SINCE 1998



Kuzmic, P. (1996) "Program DYNAFIT for the analysis of enzyme kinetic data: Application to HIV proteinase" *Anal. Biochem.* **237**, 260-273.

Kuzmic, P. (2009) "DynaFit - A software package for enzymology" *Meth. Enzymol.* **467**, 247-280.

Theoretical foundations: *Mass Action Law*

RATE IS PROPORTIONAL TO CONCENTRATION(S)

"rate" ... "derivative"

MONOMOLECULAR REACTIONS

A ⟶ products

rate is proportional to **[A]**

$-\mathrm{d}[A]/\mathrm{d}t = k[A]$

monomolecular rate constant
1 / time

BIMOLECULAR REACTIONS

A + B ⟶ products

rate is proportional to **[A] × [B]**

$-\mathrm{d}[A]/\mathrm{d}t = -\mathrm{d}[B]/\mathrm{d}t = k[A] \times [B]$

bimolecular rate constant
1 / (concentration × time)

Theoretical foundations: *Mass Conservation Law*

PRODUCTS ARE FORMED WITH THE SAME RATE AS REACTANTS DISAPPEAR

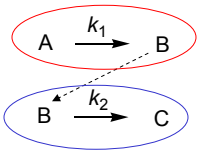
EXAMPLE

$A \longrightarrow P + Q$

$$-\frac{d[A]}{dt} = +\frac{d[P]}{dt} = +\frac{d[Q]}{dt}$$

COMPOSITION RULE ADDITIVITY OF TERMS FROM SEPARATE REACTIONS

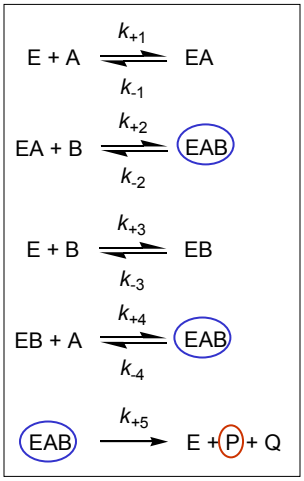
mechanism:



$$\frac{d[B]}{dt} = +k_1[A] - k_2[B]$$

Composition Rule: Example

EXAMPLE MECHANISM



RATE EQUATIONS

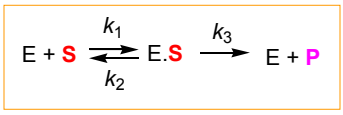
$$\frac{d[P]}{dt} = +k_{+5}[EAB]$$

$$\frac{d[EAB]}{dt} = +k_{+2}[EA]\times[B] - k_{-2}[EAB] + k_{+4}[EB]\times[A] - k_{-4}[EAB] - k_{+5}[EAB]$$

Similarly for other species...

A "Kinetic Compiler"

HOW DYNAFIT PROCESSES YOUR BIOCHEMICAL EQUATIONS



Input (plain text file):

```
E + S  ---> ES   :  k1
ES     ---> E + S :  k2
ES     ---> E + P :  k3
```

Rate terms:

$k_1 \times [E] \times [S]$

$k_2 \times [ES]$

$k_3 \times [ES]$

Rate equations:

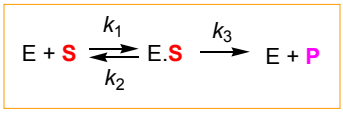
$d[\boldsymbol{E}] / dt = - k_1 \times [E] \times [S] + k_2 \times [ES] + k_3 \times [ES]$

$d[\boldsymbol{ES}] / dt = + k_1 \times [E] \times [S] - k_2 \times [ES] - k_3 \times [ES]$

Similarly for other species...

System of Simple, Simultaneous Equations

HOW DYNAFIT PROCESSES YOUR BIOCHEMICAL EQUATIONS



"The LEGO method" of deriving rate equations

Input (plain text file):

```
E + S  ---> ES   :  k1
ES     ---> E + S :  k2
ES     ---> E + P :  k3
```

Rate terms:

$k_1 \times [E] \times [S]$

$k_2 \times [ES]$

$k_3 \times [ES]$

DynaFit can analyze many types of experiments

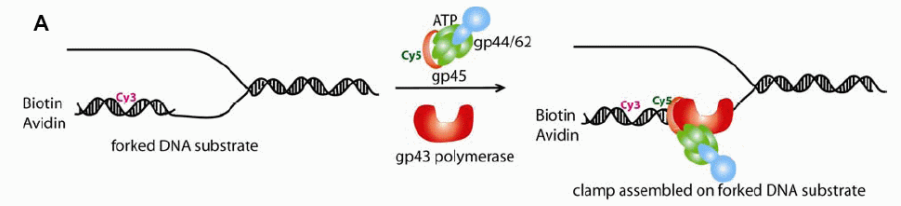
MASS ACTION LAW AND MASS CONSERVATION LAW IS APPLIED IN THE SAME WAY

chemistry
biophysics
enzymology

EXPERIMENT	DYNAFIT DERIVES A SYSTEM OF ...
Kinetics (time-course)	Ordinary differential equations (ODE)
Equilibrium binding	Nonlinear algebraic equations
Initial reaction rates	Nonlinear algebraic equations

Example: DNA + clamp / clamp loader complex

DETERMINE ASSOCIATION AND DISSOCIATION RATE CONSTANT IN AN **A + B ⇔ AB** SYSTEM



Schematic representation of the clamp loading onto forked-DNA substrate. In the forked DNA-substrate, the primer carries a Cy3 fluorescent donor and the gp45 clamp contains an acceptor Cy5 dye.

Typical email from a Ph.D. student:

[redacted] I am limited with equimolar concentrations, as I would otherwise need a lot of material for the experiments, which are really expensive because of the dyes and proteins. [redacted]

Courtesy of Senthil Perumal, Penn State University (Steven Benkovic lab)

Example: Experimental setup

ALL COMPONENTS PRESENT AT **EQUAL CONCENTRATIONS**

1. pre-mix sliding clamp (C) + clamp loader (L) to form C.L complex;
2. add DNA solution;
3. observe the formation of C.L.DNA ternary complex over time

final concentrations:

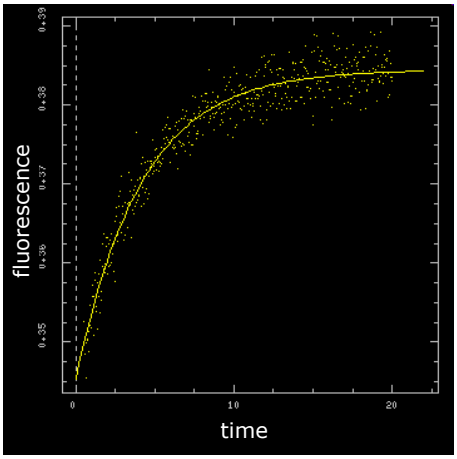
100 nM clamp ... gp45 labeled with Cy5 acceptor dye
100 nM loader ... gp44/62
100 nM DNA ... primer labeled with Cy3 donor dye

C.L complex has estimated K_d = **1 nM**, so
C.L $\rightarrow$ C+L dissociation upon adding DNA should be negligible

Courtesy of Senthil Perumal, Penn State University (Steven Benkovic lab)

Example: Raw data

JUST BECAUSE THE DATA FIT TO A MODEL DOES **NOT** MEAN THAT THE MODEL IS CORRECT!



raw fluorescence F fit to

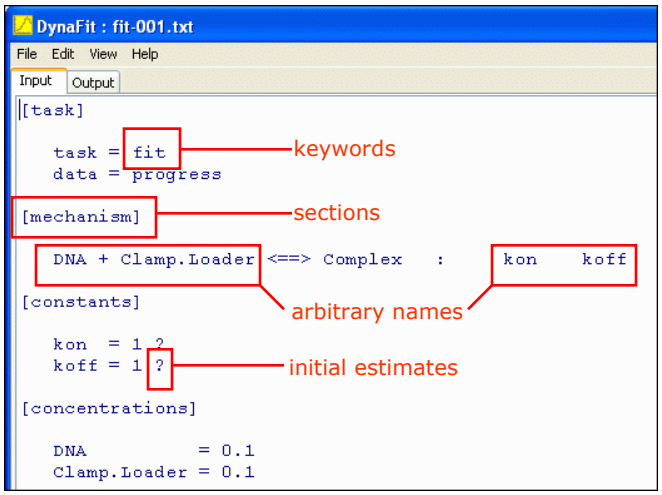
$F = A_0 + A_1 \exp(-k\,t)$

exponential model fits the data well but it is theoretically **invalid**!

Courtesy of Senthil Perumal, Penn State University (Steven Benkovic lab)

Example: Anatomy of DynaFit scripts

DYNAFIT SOFTWARE IS DRIVEN BY **TEXT "SCRIPTS"** - MINIATURE "COMPUTER PROGRAMS"



Example: DynaFit tutorial

YOUR FIRST DYNAFIT DATA-ANALYSIS SESSION

TUTORIAL

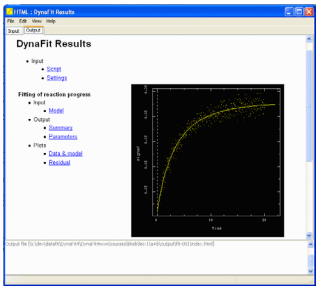
1. Start DynaFit
2. Select menu "File ... Open" or press **Ctrl**+**O**
3. Navigate to file

   ./courses/bkeb/lec-1/a+b/fit-001.txt
4. Select menu "File ... Try" or press **Ctrl**+**T**

   This is the initial estimate
5. Select menu "File ... Run" or press **Ctrl**+**U**

   Wait several seconds to finish the analysis
6. Select menu "View ... Results in External Browser"

   Navigate in the output files



Example: Detailed explanation

A BIT OF THEORY

1. Reaction order
2. Units and dimensions (scaling)
3. The DynaFit model for biomolecular kinetics
4. Initial estimates of model parameters

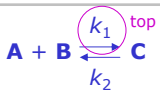
Molecularity and reaction order

IN PRACTICE WE ENCOUNTER ONLY **ZERO-**, **FIRST-**, AND **SECOND-ORDER** REACTIONS

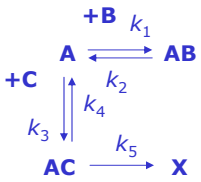
ORDER	*PHYSICAL MEANING*	*NOTATION*	*DYNAFIT NOTATION*
zero-	constant-rate influx or efflux		`X -->  : v`
first- *uni-molecular monomolecular*	isomerization or dissociation of one molecule	$A \xrightarrow{k_1} B$ $A \xrightarrow{k'_1} B + C$	`A --> B  : k1` `C --> A + B  : k1'`
second- *bimolecular*	binding (association) of two molecules	$A + B \xrightarrow{k_2} C$	`A + B --> C  : k2`

Reversible reactions and reaction mechanisms

DYNAFIT CAN HANDLE AN ARBITRARY NUMBER OF ELEMENTARY REACTIONS IN A MECHANISM

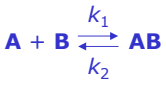
REVERSIBLE REACTION	*DYNAFIT NOTATION*
	`A + B <==> C : k1 k2` (k1: left)

MULTI-STEP MECHANISM	*DYNAFIT NOTATION*

	`A + B <==> AB : k1 k2`
	`A + C <==> AC : k3 k4`
	`AC ---> X : k5`

Dimensions of rate constants

CAREFUL ABOUT DIMENSIONS OF RATE CONSTANTS! **DIMENSIONAL ANALYSIS**



forward and reverse reaction **rates:**

$v_{\rightarrow} = k_1 [A] [B]$

$v_{\leftarrow} = k_2 [AB]$

quantity	**dimension**
v	*concentration / time*
[*X*]	*concentration*
k_1, k_2	***?***

EXAMPLE

*dimensional analysis of k_1 (**bimolecular association** rate constant)*

$$v_{\rightarrow} = k_1 [A] [B] \Rightarrow k_1 = \frac{v_{\rightarrow}}{[A] [B]} \Rightarrow \frac{\{conc.\} / \{time\}}{\{conc.\} \times \{conc.\}} = \frac{\mathbf{1}}{\mathbf{\{conc.\} \times \{time\}}}$$

Dimensions of rate and equilibrium constants

SUMMARY

$\mathbf{A + B} \underset{k_2}{\overset{k_1}{\rightleftharpoons}} \mathbf{AB}$

dissociation rate constant	k_2	**1/sec**
association rate constant	k_1	**1/(M sec)**
dissociation equilibrium constant	$K_d = k_2/k_1$	**M**
association equilibrium constant	$K_a = k_1/k_2$	**1/M**

$\mathbf{A} \underset{k_4}{\overset{k_3}{\rightleftharpoons}} \mathbf{A'}$

→ rate constant	k_3	**1/sec**
← rate constant	k_4	**1/sec**
→ equilibrium constant	$K_{\rightarrow} = k_3/k_4$	**--**
← equilibrium constant	$K_{\leftarrow} = k_4/k_3$	**--**

Example: Units (scaling) of rate constants

ALL UNITS ARE **ARBITRARY BUT MUST BE IDENTICAL** THROUGHOUT THE ENTIRE SCRIPT!

```
[mechanism]
   DNA + Clamp.Loader <==> Complex   :     kon    koff

[constants]
   kon  = 1 ?
   koff = 1 ?

[concentrations]
   DNA          = 0.1
   Clamp.Loader = 0.1

[responses]
   Complex = 1 ?

[data]
   file       ./.../d1-edit.txt
   offset     auto ?
```

= 1 **μM⁻¹s⁻¹** = 10^6 $M^{-1}.s^{-1}$

concentration units of rate constants: **μM**

100 nM

time units of rate constants: **sec**

time, s	signal, eV
0.52	0.3526
0.56	0.3484
0.60	0.3485
0.64	0.3454
0.68	0.3499
0.72	0.3502
...	

Example: The "response" coefficient

MOLAR "RESPONSE" = PROPORTIONALITY FACTOR LINKING **CONCENTRATIONS** TO **SIGNAL**

```
[mechanism]
   DNA + Clamp.Loader <==> Complex    :      kon     koff

[constants]
   kon   = 1 ?
   koff  = 1 ?

[concentrations]
   DNA           = 0.1
   Clamp.Loader  = 0.1

[responses]
   Complex = 1.00 ?

[data]
   file        ./.../d1-edit.txt
   offset      auto ?
```

one concentration unit (in this case 1 μM) of `Complex` will produce an increase in the **signal** equal to **1.00** instrument units

time, s	signal, eV
0.52	0.3526
0.56	0.3484
0.60	0.3485
0.64	0.3454
0.68	0.3499
0.72	0.3502
...	

Example: Initial estimates

NONLINEAR REGRESSION ANALYSIS ALWAYS REQUIRES **INITIAL ESTIMATES** OF THE SOLUTION

the initial estimate of rate constants

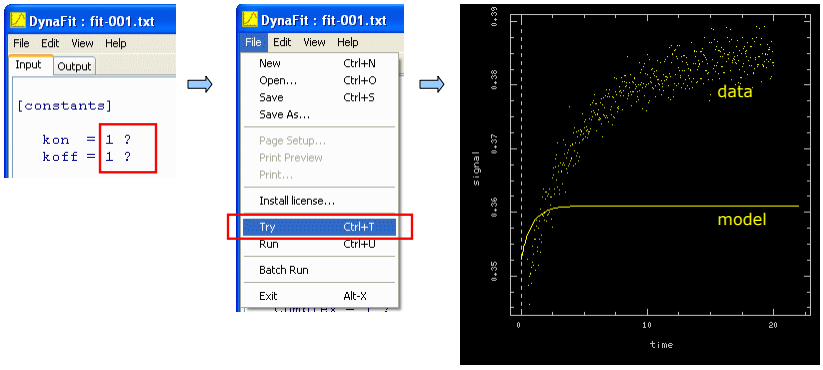
```
[mechanism]
   DNA + Clamp.Loader <==> Complex    :      kon     koff

[constants]
   kon   = 1 ?
   koff  = 1 ?

[concentrations]
   DNA           = 0.1
   Clamp.Loader  = 0.1

[responses]
   Complex = 1 ?

[data]
   file        ./.../d1-edit.txt
   offset      auto ?
```

optimized model parameters

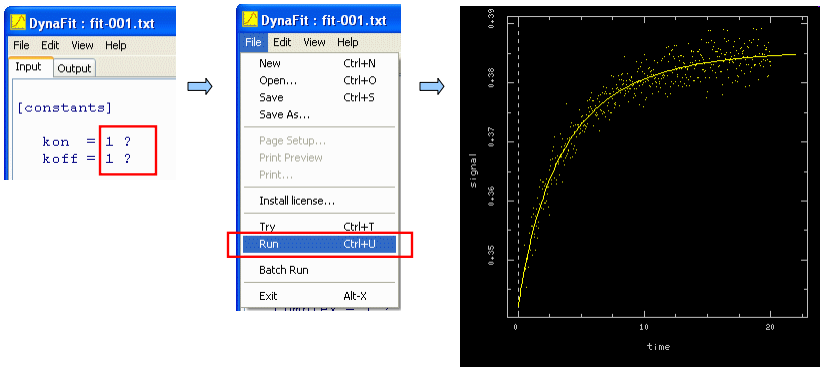
A VERY DIFFICULT PROBLEM:

HOW TO GUESS
"GOOD ENOUGH"
INITIAL ESTIMATES
OF RATE CONSTANTS?

Example: “Good” initial estimate



Example: “Good” initial estimate – results



Regression Summary

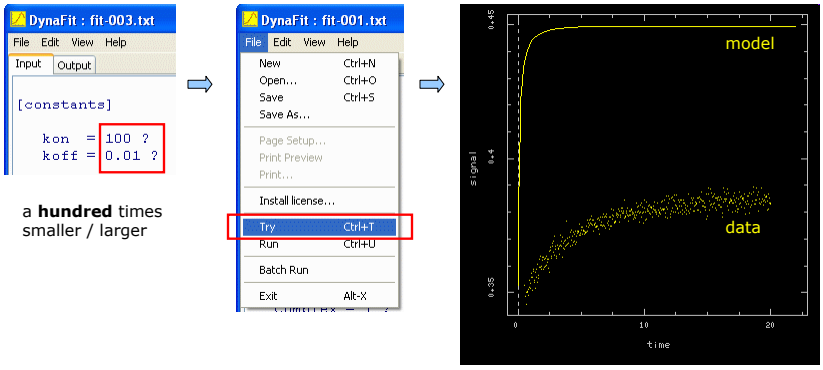
Levenberg-Marquardt Algorithm

sum of squares	0.00230769

Optimized Parameters

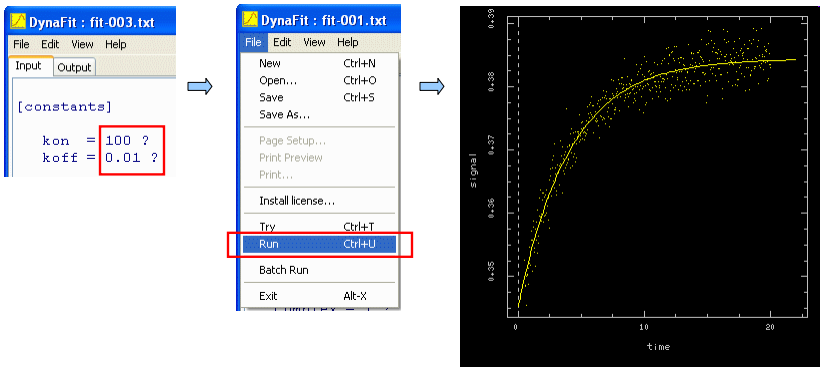
No.	Par#Set	Initial	Final	Std. Error	CV (%)	Note
#1	kon	1	2.20292	0.469206	21.30	
#2	koff	1	0.0298798	0.015383	51.48	

Example: “Bad” initial estimate



a **hundred** times smaller / larger

Example: “Bad” initial estimate – results



Regression Summary

Levenberg-Marquardt Algorithm

sum of squares	0.00235435

Optimized Parameters

No.	Par#Set	Initial	Final	Std. Error	CV (%)	Note
#1	kon	100	**0.164621**	3.38443	2055.89	
#2	koff	0.01	**0.211276**	0.609154	288.32	

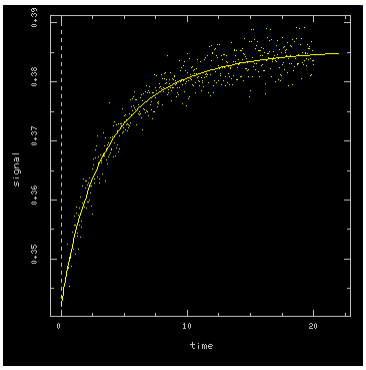
Example: "Good" vs. "Bad" results - comparison

	initial estimate	sum of squares	relative sum of sq.	"best-fit" constants	K_d, nM = k_2/k_1
"good"	$k_1 = 1$ $k_2 = 1$	0.002308	1.00	$k_1 = 2.2 \pm 0.5$ $k_2 = 0.030 \pm 0.015$	13 nM
"bad"	$k_1 = 100$ $k_2 = 0.01$	0.002354	1.02	$k_1 = 0.2 \pm 3.4$ $k_2 = 0.2 \pm 0.6$	1000 nM

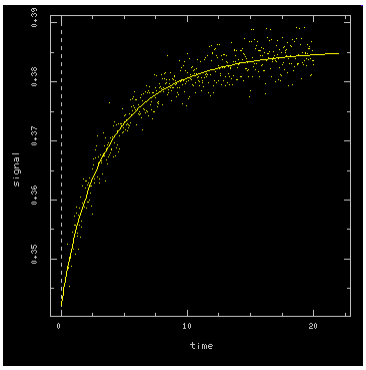
DynaFit *warnings* from running the "bad" estimate:

Error status	3
Error message	WARNING: No convergence in Lenvenberg-Marquardt algorithm. Consider increasing the number of iterations. WARNING: Hessian matrix inversion failed. This problem is ill-conditioned. Parameter errors and model inference bands are suspect. To fix this problem, try removing at least one severely redundant parameter from the fitting model.

Example: "Good" vs. "Bad" results - comparison



From "good" initial estimate



From "bad" initial estimate

not very encouraging!

Example: "Good" vs. "Bad" results - summary

1. Initial estimates "off" by a factor of **100** can produce misleading results.
2. The data/model overlay may "look good", but the results may be invalid.
3. The same applies to the residual sum of squares (only 2% difference).
4. The only indication that something went wrong might be:
   a. huge standard errors of model parameters; and
   b. various warnings from the least-squares fitter
5. The simplest possible safeguard: Use *several* different initial estimates?

   Disadvantage: how do we know which multiple estimates?

LOOKING AHEAD

DynaFit offers more reliable and convenient safeguards

- Global minimization
- Systematic combinatorial scan
- Confidence interval search

Summary and conclusions

NUMERICAL MODELS IN BIOCHEMISTRY AND BIOPHYSICS: BETTER THAN **ALGEBRAIC** EQUATIONS

1. Numerical models are applicable to all experimental conditions.
   No more "large excess of this over that".
2. Numerical models apply uniformly to all types of experiments:
   a. reaction progress (kinetics);
   b. equilibrium composition (binding);
   c. enzyme catalysis.
3. Numerical models can be automatically derived by computer.
   No more looking up algebraic equations – *if* they exist at all.
4. Main disadvantage: requirement for specialized software.
   But *DynaFit* is **free** to academic users.
5. Not a "silver bullet"! Example: the initial estimate problem.
   But this is not specific to numerical models (applies to algebraic models, too).