Evolutionary Computing

in the Study of Bio/Chemical Mechanisms

Petr Kuzmič, Ph.D.
BioKin, Ltd.

Problem:
Finding initial parameter estimates

Solution:
Differential Evolution (DE)

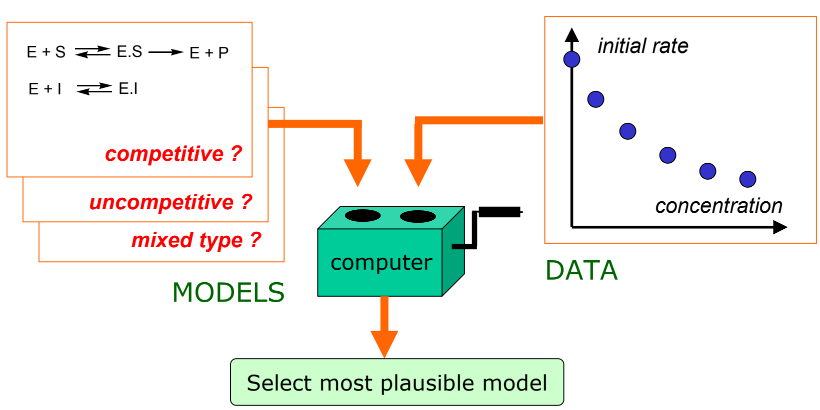
Model selection strategy:
"Supermodel Evolution"

Example:
Inhibition of Lethal Factor protease by curcumin

BioKin Ltd.

The task of mechanistic enzyme kinetics

SELECT AMONG MULTIPLE CANDIDATE MECHANISMS



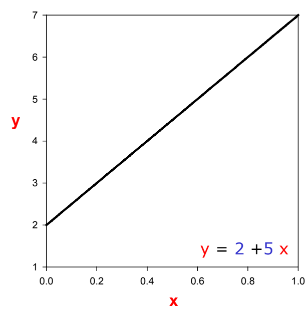
BioKin Ltd.

Most models in bio/chemical kinetics are nonlinear

LINEAR VS. NONLINEAR MODELS

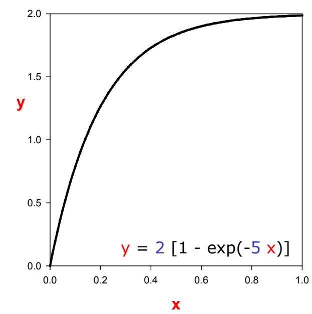
Linear

$y = \mathbf{A} + \mathbf{k}\, x$



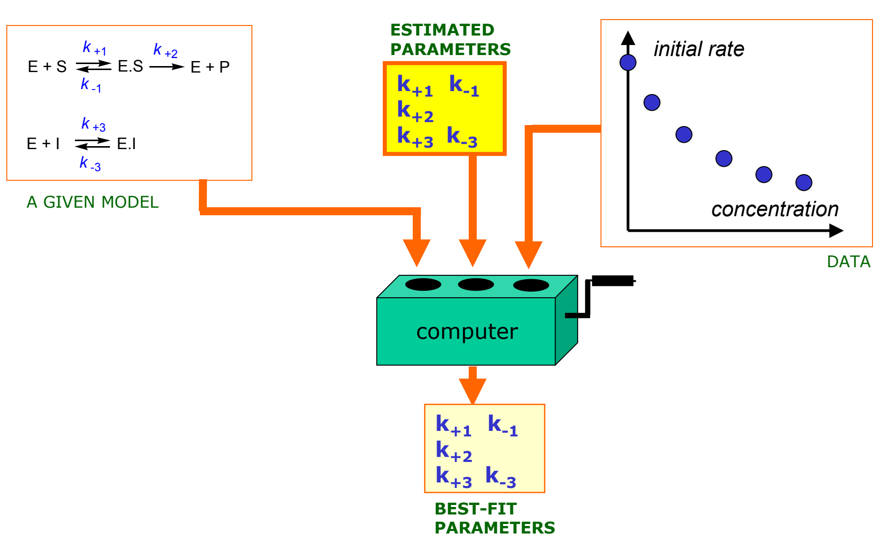
Nonlinear

$y = \mathbf{A}\, [1 - \exp(-\mathbf{k}\, x)]$



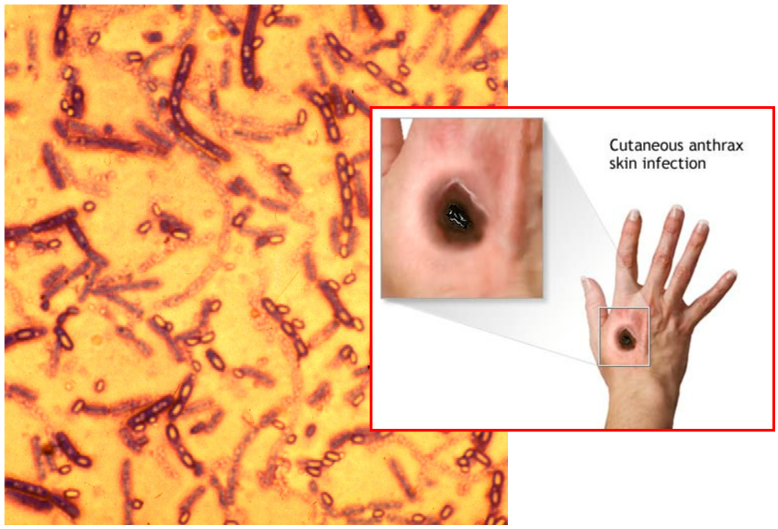
We need initial estimates of model parameters

NONLINEAR MODELS REQUIRE INITIAL ESTIMATES OF PARAMETERS



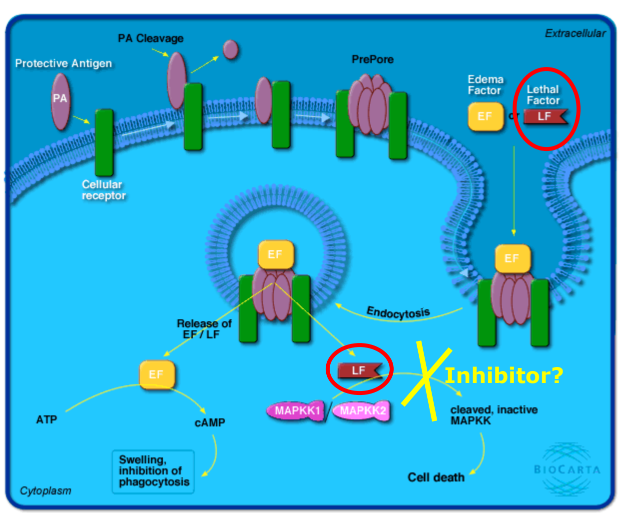
Anthrax bacillus

CUTANEOUS AND INHALATION ANTHRAX DISEASE



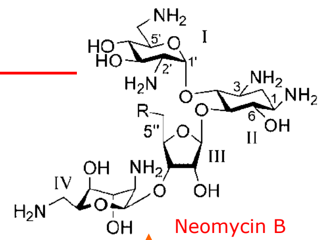
Lethal Factor (LF) protease from *B. anthracis*

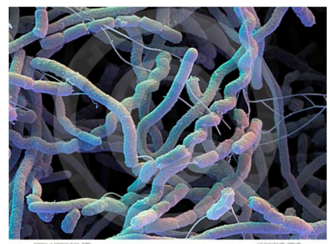
CLEAVES MITOGEN ACTIVATED PROTEIN KINASE KINASE (**MAPKK**)



Neomycin B: an aminoglycoside inhibitor

A POTENT INHIBITOR OF **LF** PROTEASE





Streptomyces fradiae

Neomycin B mechanism: mixed-type noncompetitive

NEOMYCIN B INHIBITION FOLLOWS A COMPLEX MOLECULAR MECHANISM

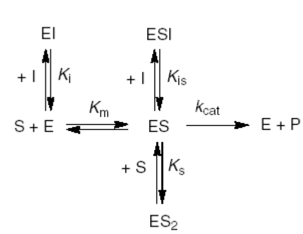
the FEBS Journal

Mixed-type noncompetitive inhibition of anthrax lethal factor protease by aminoglycosides

Petr Kuzmic[1], Lynne Cregar[2], Sherri Z. Millis[2] and Mark Goldman[2,*]

1 BioKin Ltd, Pullman, WA, USA
2 Hawaii Biotech Inc., Aiea, HI, USA

P. Kuzmic *et al.*

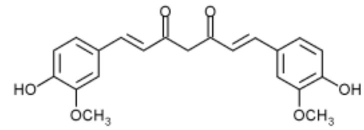


Scheme 5. Mixed-type mechanism.

Kuzmic *et al.* (2006) *FEBS J.* **273**, 3054-3062.

Curcumin: an natural product inhibitor

INHIBITION MECHANISM UNKNOWN



curcumin

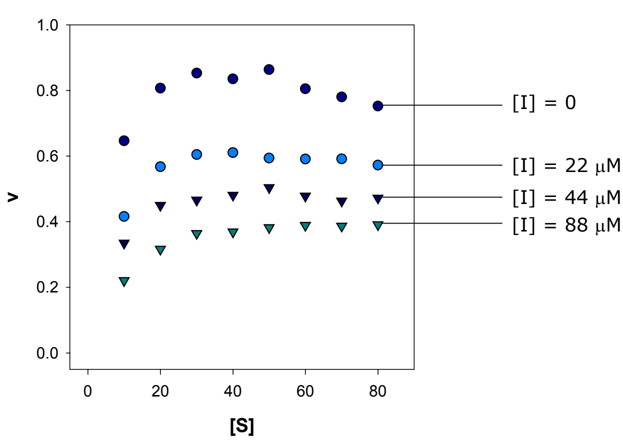
Cuminum cyminum L.

cumin

LF protease inhibition by curcumin: raw data

SUBSTRATE INHIBITION (MAXIMUM ON SUBSTRATE SATURATION CURVE)

initial reaction rate



substrate concentration

Two separate problems to solve

A PREREQUISITE FOR MODEL DISCRIMINATION = FITTING INDIVIDUAL CANDIDATE MODELS

1. **Focus on a *single* reaction mechanism:**

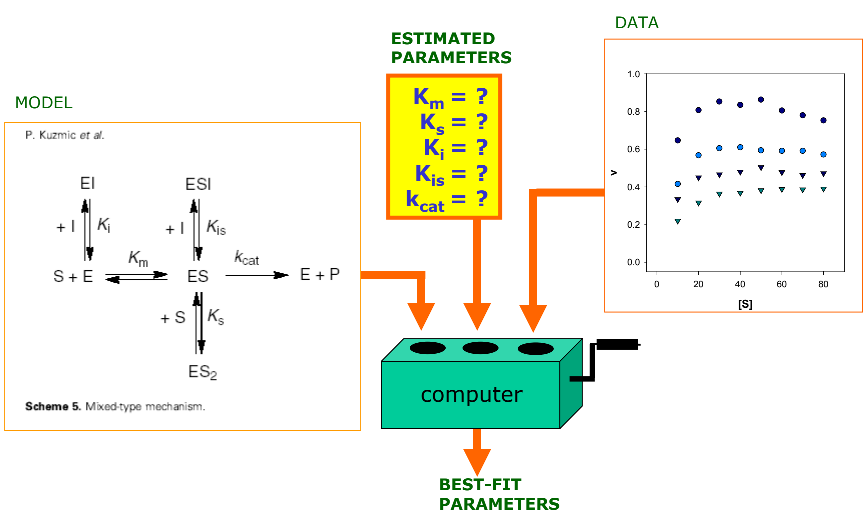
   Given a model (rate equation), find the best-fit kinetic constants

2. **Focus on *multiple* reaction mechanisms:**

   a. Repeat 1. for all candidate models (mechanisms)
   b. Select the most plausible model

The mixed-type inhibition model

CONTAINS **FIVE** KINETIC CONSTANTS

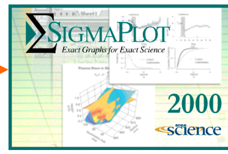


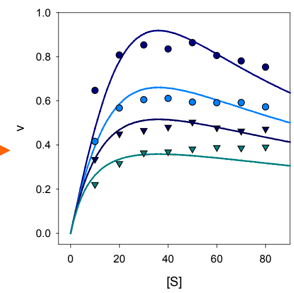
First major difficulty: Sensitivity to initial estimates

TRADITIONAL DATA-FITTING: RESULTS DEPEND ON THE INITIAL GUESS

K_m = 1
K_s = 1
K_i = 1
K_{is} = 1
k_{cat} = 1

ESTIMATE #1





K_m = 2000
K_s = 0.6
K_i ~ ∞
K_{is} = 0.0001
k_{cat} = 5100

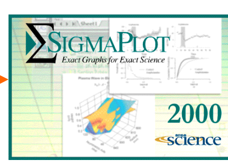
"uncompetitive" ? ← K_i

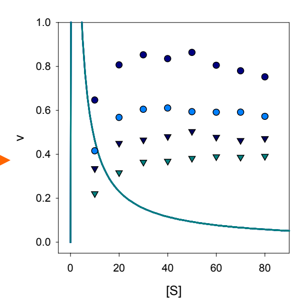
First major difficulty: Sensitivity to initial estimates

TRADITIONAL DATA-FITTING: RESULTS DEPEND ON THE INITIAL GUESS

K_m = 100
K_s = 100
K_i = 100
K_{is} = 100
k_{cat} = 100

ESTIMATE #2





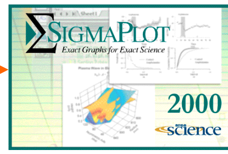
K_m < 0.000001
K_s = 0.004
K_i = 0.002
K_{is} = 0.02
k_{cat} = 54000

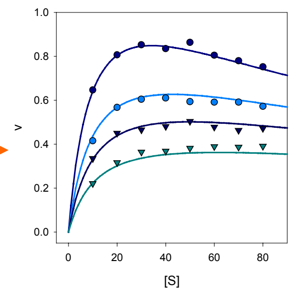
First major difficulty: Sensitivity to initial estimates

TRADITIONAL DATA-FITTING: RESULTS DEPEND ON THE INITIAL GUESS

K_m = 10
K_s = 100
K_i = 10
K_{is} = 100
k_{cat} = 100

ESTIMATE #3





K_m = 10
K_s = 110
K_i = 40
K_{is} = 14
k_{cat} = 68

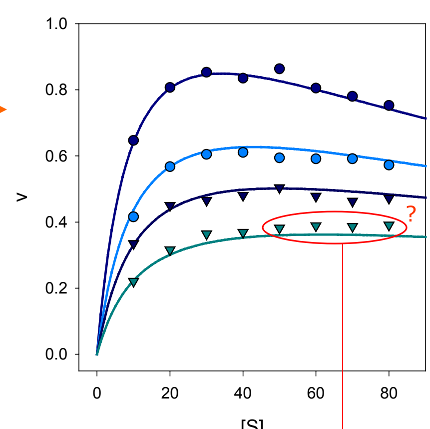
mixed-type ?

First major difficulty: Sensitivity to initial estimates

TRADITIONAL DATA-FITTING: RESULTS DEPEND ON THE INITIAL GUESS

K_m = 10
K_s = 100
K_i = 10
K_{is} = 100
k_{cat} = 100

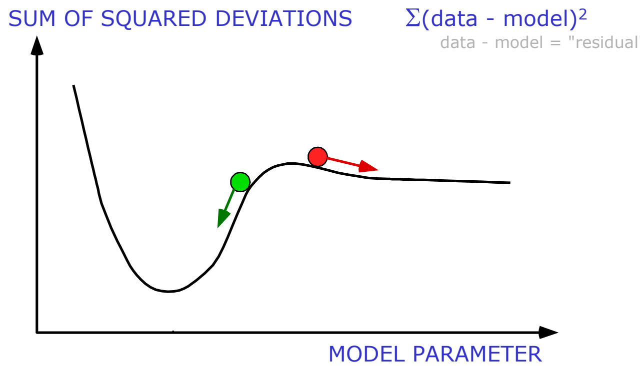
ESTIMATE #3



K_m = 10
K_s = 110
K_i = 40
K_{is} = 14
k_{cat} = 68

Is this the <u>best</u> we can do with the mixed-type model?

The crux of the problem: Finding *global* minima

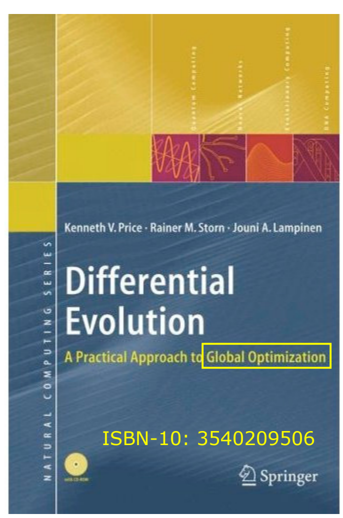


- Least-squares fitting **only** goes "downhill"
- **How do we know where to start?**

Charles Darwin to the rescue

BIOLOGICAL EVOLUTION IMITATED IN "DE"

Charles Darwin (1809-1882)



Biological metaphor: "Gene, allele"

BIOLOGY	COMPUTER
gene	• sequence of bits representing a number
...AAGTCG...GTAACCGG... "keratin"	...0111001100000110110110011... "K_M" "k_{cat}"
four-letter alphabet variable length	• two letter alphabet • fixed length (16 or 32 bits)

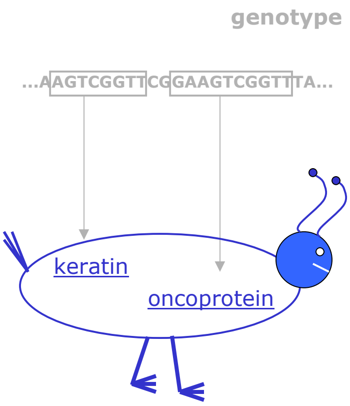
"Chromosome, genotype, phenotype"

BIOLOGY	COMPUTER
genotype	• particular combination of all model parameters
...AAGTCGGTTCGGAAGTCGGTTTA...	011010110110 011110011010 001111101101
keratin, oncoprotein	V_{max}=1.23, K_M=4.56, K_{is}=78.9
phenotype	**full set of parameters**

$$v = V_{max} \frac{[S]/K_M}{1 + [S]/K_M + [S]^2/K_M K_{is}}$$

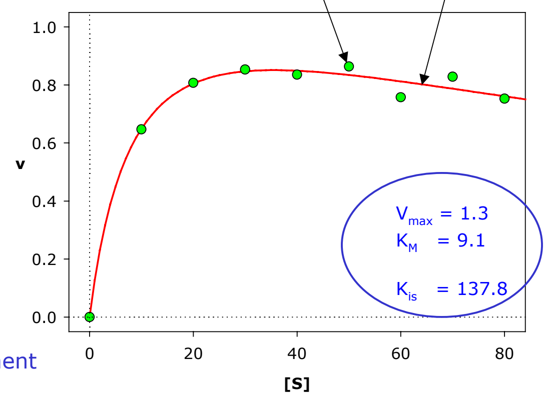
"Organism, fitness"

BIOLOGY | COMPUTER



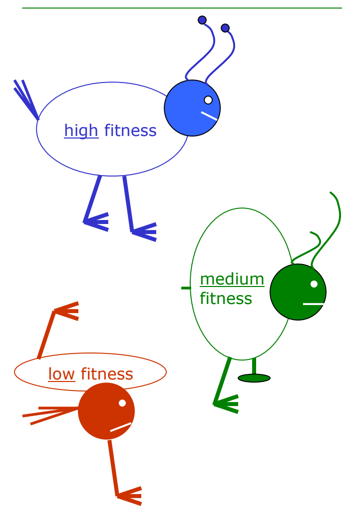
FITNESS:
"agreement" with the environment

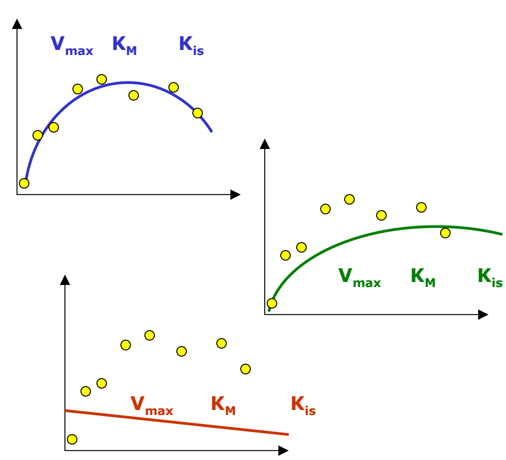
• FITNESS:
agreement between the data and the model



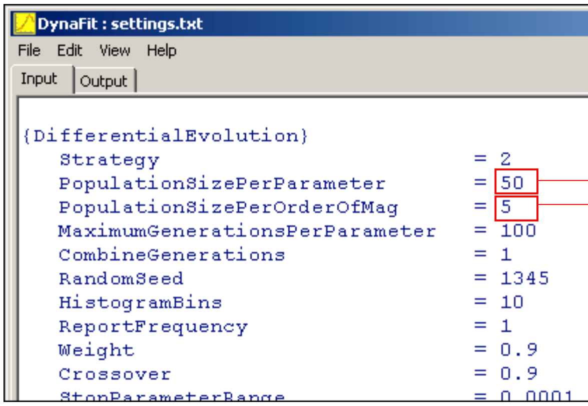
"Population"

BIOLOGY | COMPUTER





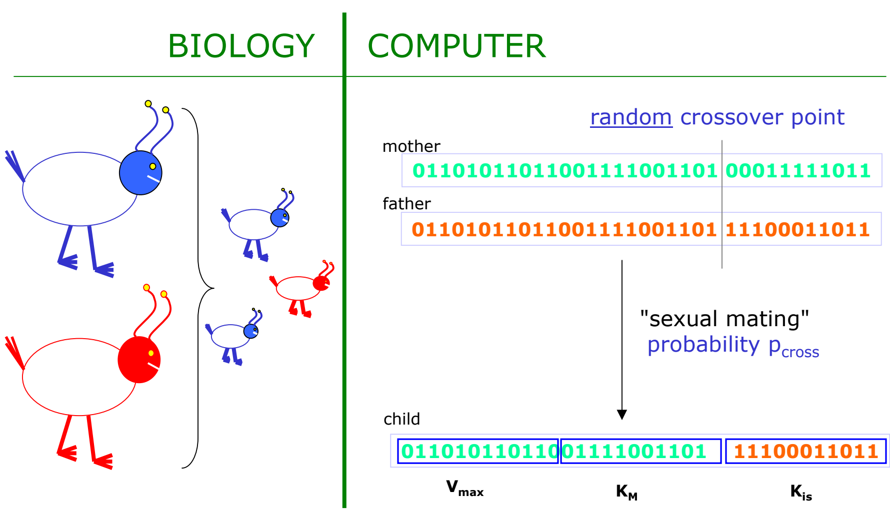
DE Population size in DynaFit



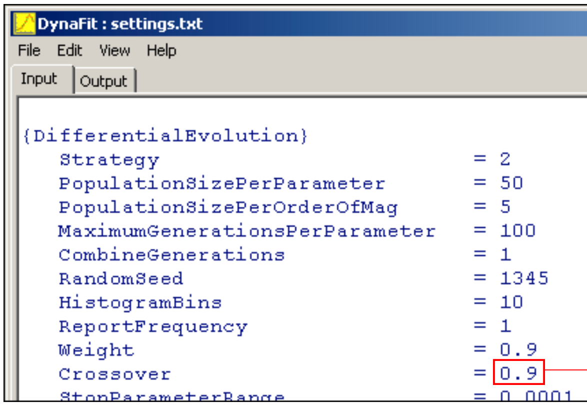
number of population members **per optimized model parameter**

number of population members **per order of magnitude**

"Sexual reproduction, crossover"



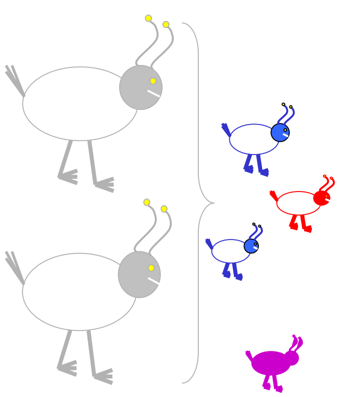
DE Crossover probability in DynaFit

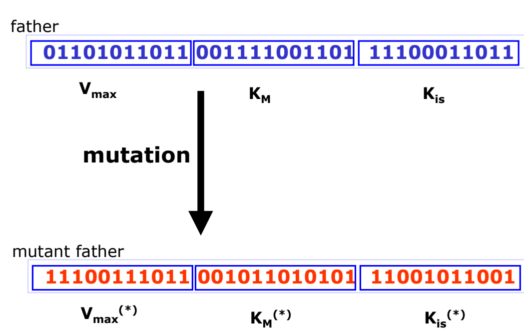


probability that ***child*** inherits ***father's*** genes, not ***mother's*** genes

"Mutation, genetic diversity"

BIOLOGY	COMPUTER





"Mutation, genetic diversity"

THE "**DIFFERENTIAL**" IN DIFFERENTIAL EVOLUTION ALGORITHM - **STEP 1**

Compute difference between two randomly chosen "uncle" phenotypes

uncle#1

01101011011	001111001101	11100011011
$V_{max}^{(1)}$	$K_M^{(1)}$	$K_{is}^{(1)}$

subtract

uncle#2

11100111011	001011010101	11001011001
$V_{max}^{(2)}$	$K_M^{(2)}$	$K_{is}^{(2)}$

↓

uncle#2 **minus** uncle #1

11100111011	001011010101	11001011001
$V_{max}^{(2)}-V_{max}^{(1)}$	$K_M^{(2)}-K_M^{(1)}$	$K_{is}^{(2)}-K_{is}^{(1)}$

"Mutation, genetic diversity"

THE "**DIFFERENTIAL**" IN DIFFERENTIAL EVOLUTION ALGORITHM - **STEP 2**

Add **weighted** difference between two "uncle" phenotypes to "father"

father

01101011011	001111001101	11100011011
V_{max}	K_M	K_{is}

add a <u>fraction</u> of

uncle#2 **minus** uncle #1

11100111011	001011010101	11001011001
$V_{max}^{(2)}-V_{max}^{(2)}$	$K_M^{(2)}-K_M^{(1)}$	$K_{is}^{(2)}-K_{is}^{(1)}$

↓

mutant father

11100111011	001011010101	11001011001
V_{max}^{*}	K_M^{*}	K_{is}^{*}

"Mutation, genetic diversity"

THE "**DIFFERENTIAL**" IN DIFFERENTIAL EVOLUTION ALGORITHM

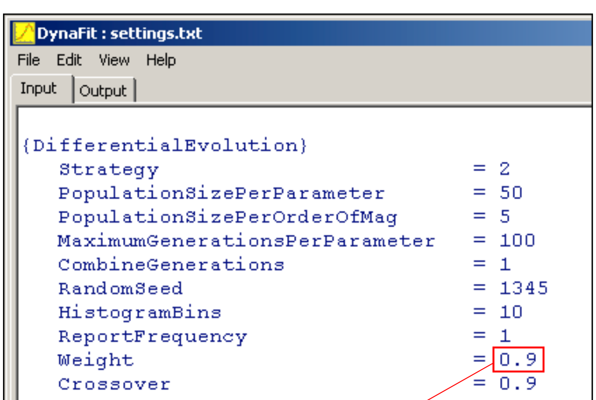
EXAMPLE: Michaelis-Menten equation $\mathrm{v} = V_{max} \dfrac{[S]}{[S] + K_M}$

"father" "uncle 1" "uncle 2"

"mutant father" → $K_M^{*} = K_M + \mathbf{F} \times (K_M^{(1)} - K_M^{(2)})$

weight (fraction)
mutation rate

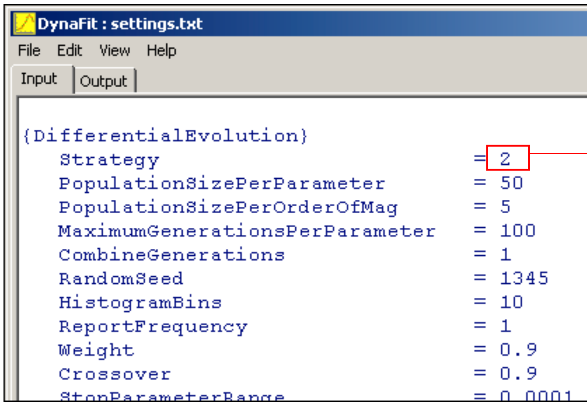
DE Mutation rate in DynaFit



fractional difference used in mutations

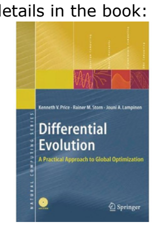
$K_M^{*} = K_M + \mathbf{F} \times (K_M^{(1)} - K_M^{(2)})$

DE Mutation strategies in DynaFit



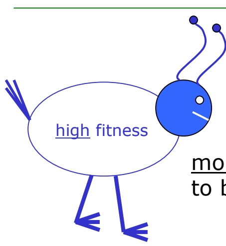
six different mutation strategies (1, 2, ... 6)

details in the book:



"Selection"

BIOLOGY | COMPUTER



more likely to breed

low sum of squares

01101011011	00111100110	00011111011
V_{max}	K_M	K_{is}

more likely to be carried to the next generation

less likely to breed

high sum of squares

00000000001	11111111111	00000000000
V_{max}	K_M	K_{is}

less likely to be carried to the next generation

Basic Differential Evolution Algorithm - Summary

1 **Randomly** create the initial population (size **N**)

Repeat until almost all population members have very high fitness:

2 Evaluate fitness: sum of squares for all population members

3 Mutation: **random** gene modification (mutate *father*, weight **F**)

4 Sexual reproduction: **random** crossover with probability $\mathbf{P_{cross}}$

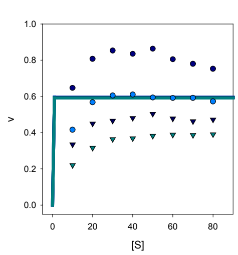
5 Natural selection: keep *child* in gene pool if more fit than *mother*

Application to curcumin: Mixed-type mechanism

THREE EXAMPLES OF POPULATION MEMBERS (**POPULATION SIZE *n* = 845**)

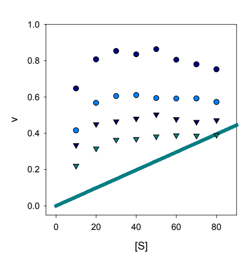
POPUL. MEMBER #1

Km	3.3355e-11
Ks	7868500
Ki	106.99
Kis	4.0737e-9
kcat	29.783



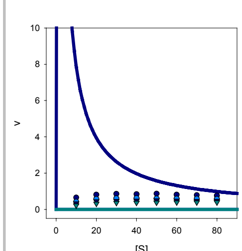
POPUL. MEMBER #7

Km	3713600
Ks	60218
Ki	427880
Kis	153.55
kcat	918170



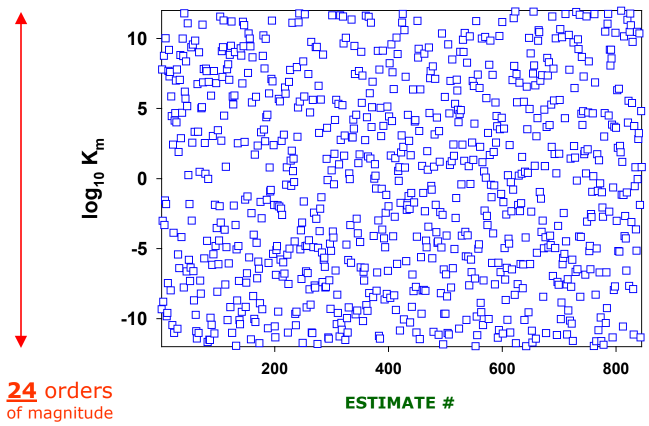
POPUL. MEMBER #642

Km	65.7630
Ks	1.0488e-3
Ki	8.7928e-10
Kis	8.1374e+20
kcat	3760400



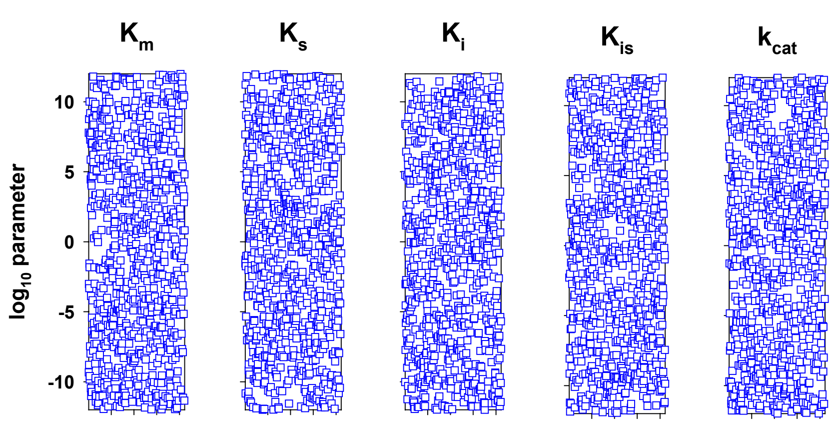
Application to curcumin: Mixed-type mechanism

INITIAL DISTRIBUTION OF THE **MICHAELIS-CONSTANT K_M**



Application to curcumin: Mixed-type mechanism

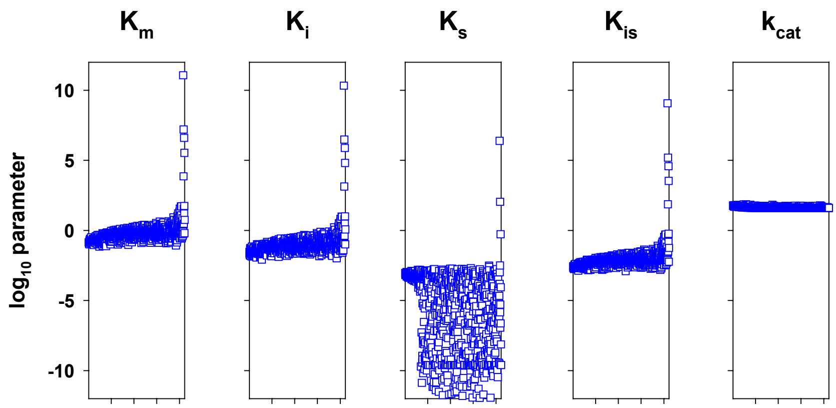
INITIAL DISTRIBUTION OF ALL MODEL PARAMETERS



Generation #1

Application to curcumin: Mixed-type mechanism

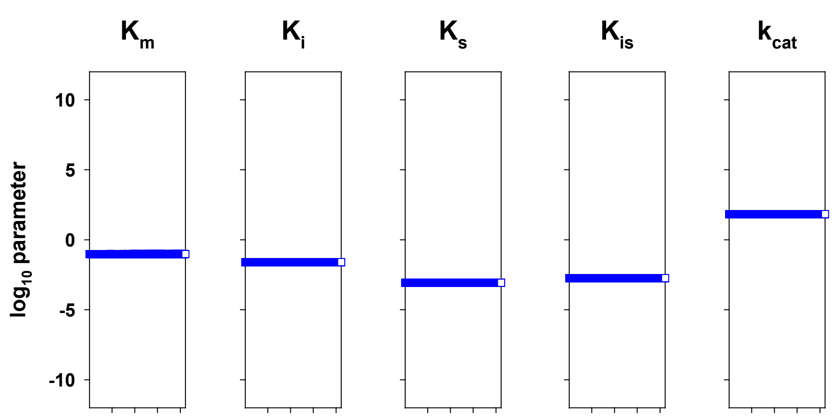
INTERMEDIATE DISTRIBUTION OF MODEL PARAMETERS (**EVOLUTION IS ONGOING**)



Generation #20

Application to curcumin: Mixed-type mechanism

FINAL DISTRIBUTION OF MODEL PARAMETERS (**EVOLUTION IS COMPLETED**)

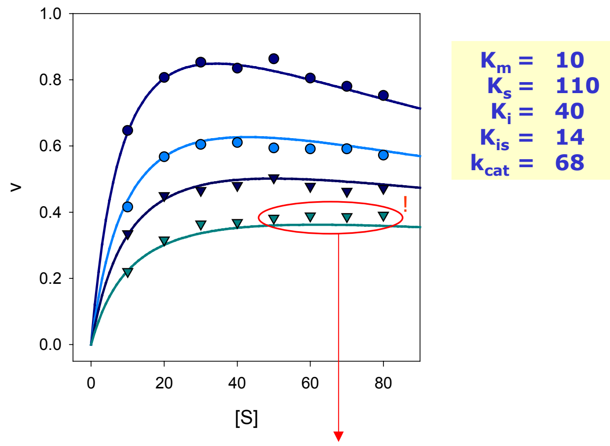


Generation #125

Application to curcumin: Mixed-type mechanism

THE SUPER-ORGANISM

The fittest "phenotype" in generation #125



Yes, this **is** the best we can do with the mixed-type model!

The model selection problem remains

A PREREQUISITE FOR MODEL DISCRIMINATION = FITTING INDIVIDUAL CANDIDATE MODELS

1. **Focus on a *single* reaction mechanism:** ✓

   Given a model (rate equation), find the best-fit kinetic constants

2. **Focus on *multiple* reaction mechanisms:**

   a. Repeat 1. for **all candidate models** (mechanisms)
   b. Select the most plausible model

?

Model proliferation: CYP450 / reductase example

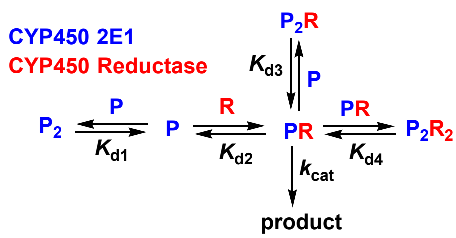
COMBINATORIAL PROLIFERATION OF POSSIBLE MECHANISMS

EXAMPLE

Global Analysis of Protein-Protein Interactions Reveals
Multiple Cytochrome P450 2E1 / Reductase Complexes

Arvind P. Jamakhandi[‡], Petr Kuzmič[§], Daniel E. Sanders[‡], and Grover P. Miller[‡]

Journal: ***Biochemistry*** Ms# **BI-2006-003476.R2** IN PRESS



Model proliferation: CYP450 / reductase example

COMBINATORIAL PROLIFERATION OF POSSIBLE MECHANISMS

P = cytochrome P450 (2E1)
R = cytochrome reductase

	Complex				
Model#	**PR**	**P2R**	**PR2**	**P2R2**	**P2**
1	A				
2	A	A			
3	A		A		
4	A	A	A		
5	A			A	
6	A	A		A	
7	A		A	A	
8	A	A	A	A	N
9	A				N
10	A	A			N
11	A		A		N
12	A	A	A		N
13	A			A	N
14	A	A		A	N
15	A		A	A	N
16	A	A	A	A	N

A = catalytically active
N = inactive

42 separate mechanisms were examined

The "Supermodel" approach

CREATE AN **AGGREGATE** MODEL ENCOMPASSING ALL POSSIBLE INTERACTIONS

	Complex				
Model#	PR	P2R	PR2	P2R2	P2
1	A				
2	A	A			
3	A		A		
4	A	A	A		
5	A			A	
6	A	A		A	
7	A		A	A	
8	A	A	A	A	N
9	A				N
10	A	A			N
11	A		A		N
12	A	A	A		N
13	A			A	N
14	A	A		A	N
15	A		A	A	N
16	A	A	A	A	N

the most complex (realistic) model

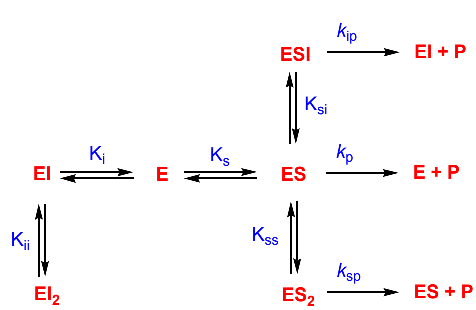
1. Consider only the **most complex** model
2. Evolve all parameters using "DE"
3. Identify redundant parameters
   by examining the final distribution of fittest phenotypes
4. Eliminate redundant parameters
   thereby reducing the model ("small is beautiful")

The "Supermodel" for LF inhibition by curcumin

COMPILE AN AGGREGATE OF ALL POSSIBLE MOLECULAR INTERACTIONS

ASSUMPTIONS

- Substrate can bind with 1:1 or 2:1 stoichiometry
- Inhibitor can bind with 1:1 or 2:1 stoichiometry
- Substrate and inhibitor can bind at the same time
- Any enzyme-substrate complex can have catalytic activity

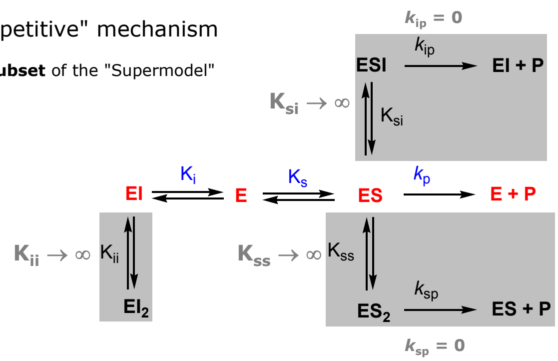


The "Supermodel" includes all standard mechanisms

STANDARD MECHANISMS DIFFER ONLY BY VALUES OF KINETIC CONSTANTS IN THE "SUPERMODEL"

"Competitive" mechanism

is a **subset** of the "Supermodel"

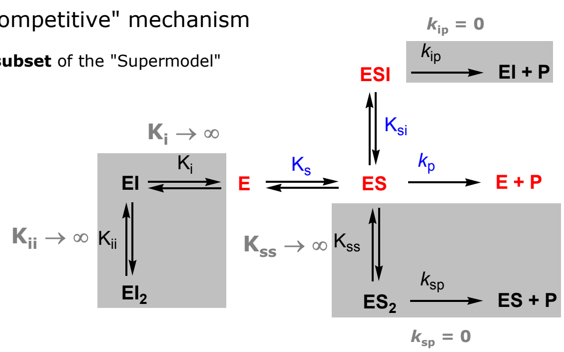


The "Supermodel" includes all standard mechanisms

STANDARD MECHANISMS DIFFER ONLY BY VALUES OF KINETIC CONSTANTS IN THE "SUPERMODEL"

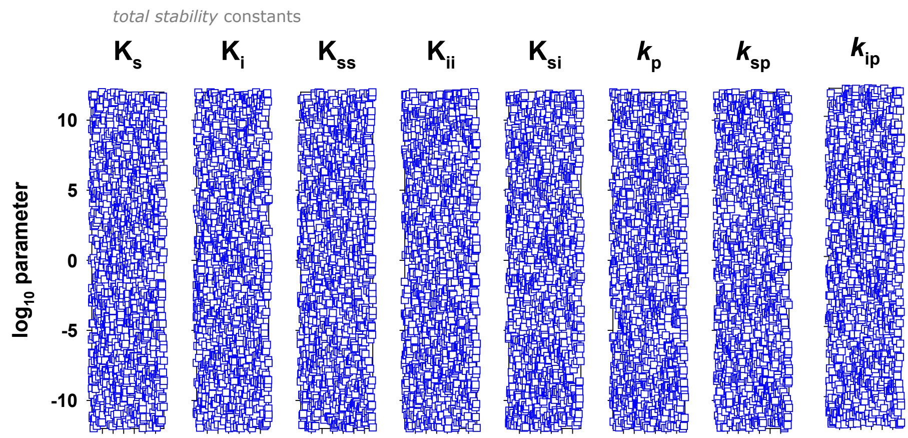
"Uncompetitive" mechanism

is a **subset** of the "Supermodel"



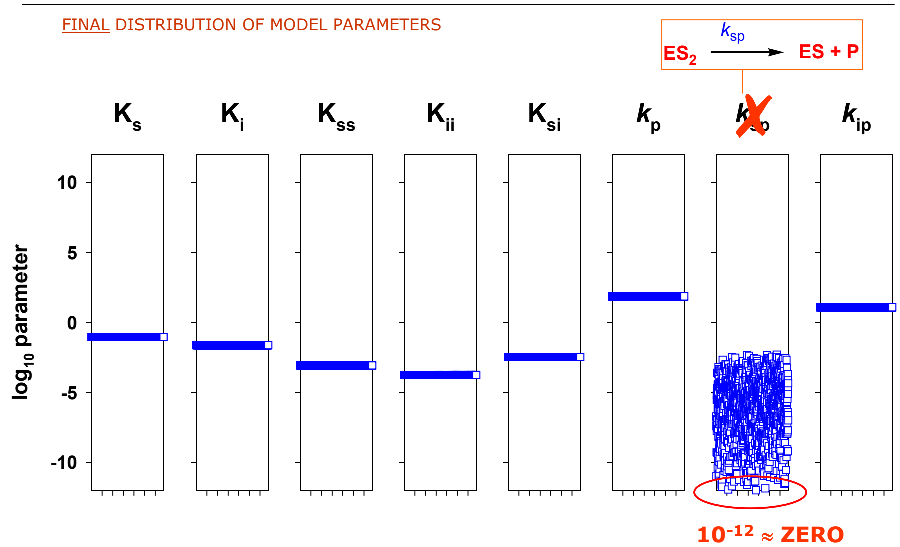
"Supermodel" evolution: Curcumin inhibition of LF

INITIAL DISTRIBUTION OF MODEL PARAMETERS (POPULATION SIZE = **1355**)



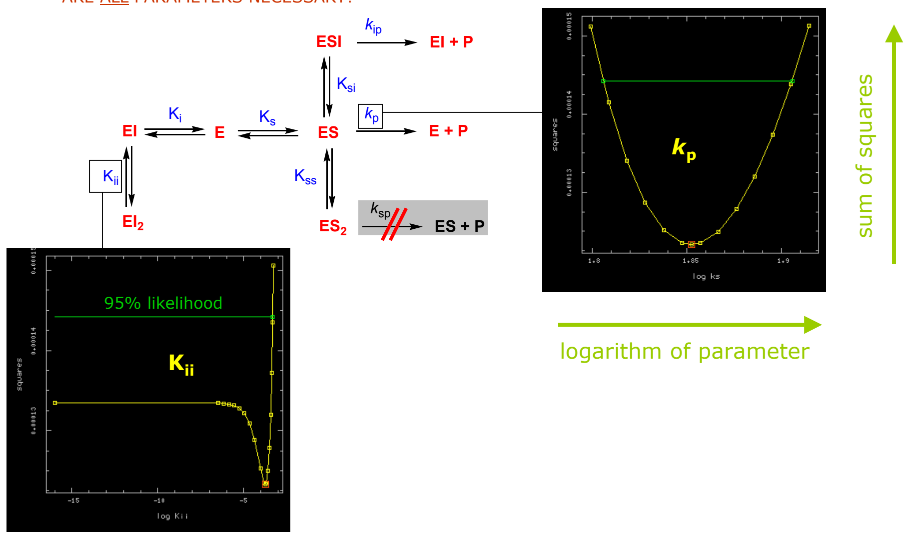
"Supermodel" evolution: Curcumin inhibition of LF

FINAL DISTRIBUTION OF MODEL PARAMETERS



Curcumin inhibition of LF: Confidence intervals

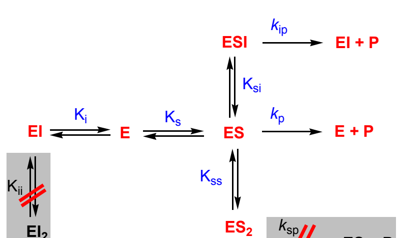
ARE ALL PARAMETERS NECESSARY?

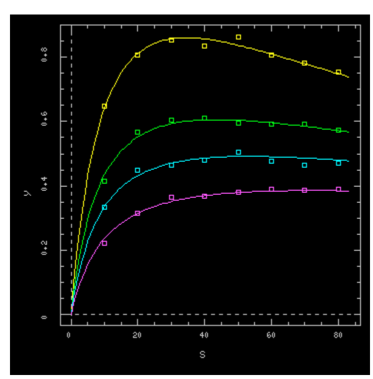


BioKin Ltd.

Curcumin inhibition of LF: Final model

FULLY EVOLVED "SUPERMODEL"





Parameter	Best Fit	Low	High
Ks	0.091	0.071	0.118
Ki	0.030	0.023	0.039
Kss	0.00085	0.00077	0.00094
Ksi	0.0029	0.0024	0.0036
*k*s	70.5	63.4	79.4
*k*si	9.2	6.7	11.4

"Partial Mixed-Type Noncompetitive"
"With Substrate Inhibition"

BioKin Ltd.

Summary and Conclusions

DIFFERENTIAL EVOLUTION HAS PROVED USEFUL IN BIO/CHEMICAL KINETICS

- Given a particular model,
  we can always find **globally optimal** parameters:

DE avoids false minima and other pathologies in nonlinear data fitting

- Given a set of possible bio/chemical processes,
  we can **"evolve" the most plausible model:**

- 1. Evolve a "Supermodel"
- 2. Exclude redundant reaction steps
- 3. Check confidence intervals

Acknowledgements: Lethal Factor protease work