Optimal Experiment Design

for Dose-Response Screening of Enzyme Inhibitors

Petr Kuzmic, Ph.D.
BioKin, Ltd.
WATERTOWN, MASSACHUSETTS, U.S.A.

PROBLEM

- Most assays in a typical screening program are *not informative*

SOLUTION

- Abandon "batch design" of dose-response experiments
- Use "sequential design" based on *D-Optimal Design Theory*

- *Save 50%* of screening time, labor, and material resources

BioKin

Two basic types of experiments

BATCH VS. SEQUENTIAL DESIGN OF ANY RESEARCH PROJECT

design = choice of screening **concentrations**

BATCH DESIGN OF EXPERIMENTS

1. decide beforehand on the design of a ***complete series of experiments***
2. perform all experiments in the series ***without analyzing interim*** *results*
3. ***analyze entire batch*** of accumulated data
4. issue final report

SEQUENTIAL DESIGN OF EXPERIMENTS

1. decide on the design of only ***one*** (or a small number of) ***experiment***(s)
2. perform one experiment
3. analyze interim results; did we accumulate enough experiments?
4. if not, ***go back to step 1***, otherwise ...
5. issue final report

Analogy with clinical trials

ADAPTIVE CLINICAL TRIALS (ACT): ADJUST THE EXPERIMENT DESIGN AS TIME GOES ON

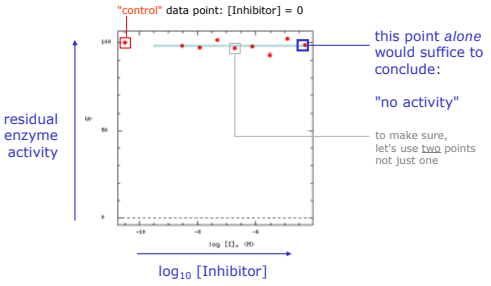
Borfitz, D.: "Adaptive Designs in the Real World" *BioIT World*, June 2008

- assortment of statistical approaches including "early stopping" and "dose-finding"
- interim data analysis
- reducing development timelines and costs by utilizing actionable information sooner
- *experts*: **Donald Berry**, chairman of the Department of Biostatistics University of Texas MD Anderson Cancer Center
- *software vendors*: **Cytel, Tessela**
- *industry pioneers*: **Wyeth** 1997 "Learn and Confirm" model of drug development

"slow but sure restyling of the research enterprise"

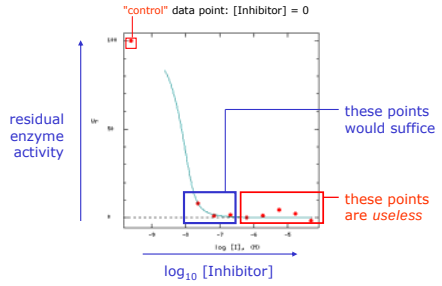
What is wrong with this dose-response curve?

THE "RESPONSE" IS INDEPENDENT OF "DOSE": NOTHING LEARNED FROM MOST DATA POINTS



What is wrong with *this* dose-response curve?

THE SAME STORY: MOST DATA POINTS ARE USELESS



Why worry about doing useless experiments?

IN CASE THE REASONS ARE NOT OBVIOUS:

Academia:	• time	Publish your paper on time for grant renewal.
	• money	Spend less on chemicals, hire a post-doc.
	• fame	Invent a drug, get the Nobel Prize.
Industry:	• time	Beat "the competition" to market.
	• money	Spend less on chemicals, hire a post-doc.
	• security	Invent a drug, avoid closure of Corporate R&D.

On a more serious note...

THERE ARE **VERY GOOD** REASONS TO GET SCREENING PROJECTS DONE AS QUICKLY AS POSSIBLE

guardian.co.uk

News | Sport | Comment | Culture | Business | Money | Life & style

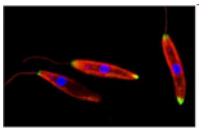
News > Society > International aid and development

University hunts cure for parasitic infections

Tim Radford
The Guardian, Wednesday October 26 2005
Article history

Six Scottish scientists have been given £13m to find a drug to treat three of the world most unpleasant diseases. Sleeping sickness, leishmaniasis and Chagas' disease are all caused by parasites

The next step is to build a library of 100,000 compounds, and begin testing to see which might afflict the parasites without harming the host. In effect, university scientists will begin the sifting process normally funded by pharmaceutical companies. The aim is to have within five years something that the World Health Organisation, or a drug manufacturer, or one of the big charities, could develop.



Leishmania major

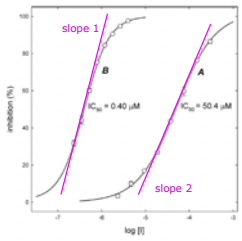
Photo: E. Dráberová
Academy of Sciences of the Czech Republic

Theoretical foundations: The inhibition constant

DO NOT USE IC_{50}. THE INHIBITION CONSTANT IS MORE INFORMATIVE

Kuzmič *et al.* (2003) *Anal. Biochem.* **319**, 272–279

$E + I \rightleftharpoons E{\bullet}I$ $\quad K_i = [E]_{eq}[I]_{eq} / [E.I]_{eq}$ $\quad K_i$ ... equilibrium constant



"Morrison equation" ✓

$$v = v_0 \frac{[E]_0 - [I]_0 - K_T + \sqrt{([E]_0 - [I]_0 - K_T)^2 + 4[E]_0 K_T}}{2[E]_0}$$

Four-parameter logistic equation

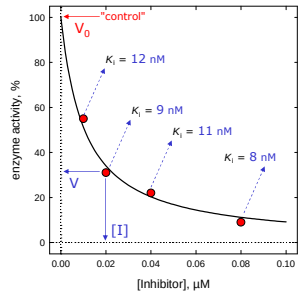
$$p = p_{min} + \frac{p_{max} - p_{min}}{1 + ([I]_0 / I_{50})^n}$$ "Hill slope"

no clear physical meaning !

Theoretical foundations: The "single-point" method

AN **APPROXIMATE** VALUE OF THE INHIBITION CONSTANT FROM A **SINGLE** DATA POINT

Kuzmič *et al.* (2000) *Anal. Biochem.* **281**, 62–67



Relative rate

$V_r = V/V_0$

Single-point formula:

$$K_i = \frac{[I] - [E](V_r - 1)}{1/V_r - 1}$$

Theoretical foundations: *Optimal Design Theory*

NOT ALL POSSIBLE EXPERIMENTS ARE EQUALLY INFORMATIVE

BOOKS:

- Fedorov (1972) "Theory of Optimal Experiments"
- Atkinson & Donev (1992) "Optimum Experimental Designs"

EDITED BOOKS:

- Endrényi (Ed.) (1981) "Kinetic Data Analysis: **Design** and Analysis of Enzyme and Pharmacokinetic Experiments"
- Atkinson *et al.* (Eds.) (2000) "Optimum Design 2000"

JOURNAL ARTICLES:

- **Thousands** of articles in many journals.
- Several articles deal with experiments in **enzymology** / **pharmacology**.

Optimal design of ligand-binding experiments

SIMPLE LIGAND BINDING AND HYPERBOLIC SATURATION CURVES

Endrényi & Chang (1981) *J. Theor. Biol.* **90**, 241-263

SUMMARY:

- Protein (P) binding with ligand (L)

dissociation constant

$P + L \rightleftharpoons^{K_d} P{\bullet}L$

- Vary *total* ligand concentration [*L*]
  Observe *bound* ligand concentration [L_B]
- Fit data to nonlinear model:

$$[L_B] = \frac{1}{2}\left([L] + K_d + [P] + \sqrt{([L] + K_d + [P])^2 - 4[P][L]}\right)$$

TWO OPTIMAL LIGAND CONCENTRATIONS (we need at least **two** data points):

$[L]_1 = [L]_{max}$ ---------- maximum **feasible** [Ligand]

$$[L]_2 = (K_d + [P]) \frac{+(K_d - [P])[L]_1 - \left(\sqrt{([L]_1 + K_d + [P])^2 - 4[P][L]_1} - K_d - [P]\right)(K_d + [P])}{-(K_d + [P])[L]_1 + \left(\sqrt{([L]_1 + K_d + [P])^2 - 4[P][L]_1} - K_d - [P]\right)(K_d - [P])}$$

Optimal design of enzyme inhibition experiments

THIS TREATMENT APPLIES BOTH TO "TIGHT BINDING" AND "CLASSICAL" INHIBITORS

Kuzmič (2008) *manuscript in preparation*

SUMMARY:

- Enzyme (E) binding with inhibitor (I)

dissociation constant

$E + I \rightleftharpoons^{K_i} E{\bullet}I$

- Vary total inhibitor concentration [*I*]
  Observe residual enzyme activity, proportional to $[E]_{free}$
- Fit data to nonlinear model:

$$V = V_0 \frac{1}{2[E]}\left([E] - [I] - K_i + \sqrt{([E] - [I] - K_i)^2 + 4K_i[E]}\right)$$ "Morrison Equation"

TWO OPTIMAL INHIBITOR CONCENTRATIONS (we need at least **two** data points):

$[I]_1 = 0$ ---------- **control** experiment (zero inhibitor)

$[I]_2 = K_i + [E]$

A problem with optimal design for *nonlinear* models

A CLASSIC **CHICKEN & EGG** PROBLEM

Endrényi & Chang (1981) *J. Theor. Biol.* **90**, 241-263 **PROTEIN/LIGAND BINDING**

$$[L]_1 = [L]_{\max}$$

$$[L]_2 = \frac{+(K_d - [P])[L]_1 - \left(\sqrt{([L]_1 + K_d + [P])^2 - 4[P][L]_1} - K_d - [P]\right)(K_d + [P])}{-(K_d + [P])[L]_1 + \left(\sqrt{([L]_1 + K_d + [P])^2 - 4[P][L]_1} - K_d - [P]\right)(K_d - [P])}$$

Kuzmič (2008) *manuscript in preparation* **ENZYME INHIBITION**

$$[I]_1 = 0$$

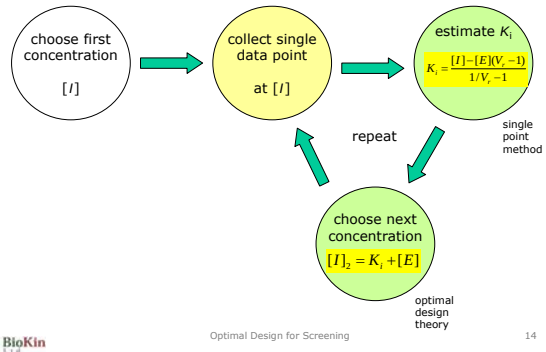
$$[I]_2 = K_i + [E]$$

a model parameter ("final answer") we are trying to determine by the experiment ... being planned!

We must guess the answer before we begin designing the experiment.

A solution for designed enzyme inhibition studies

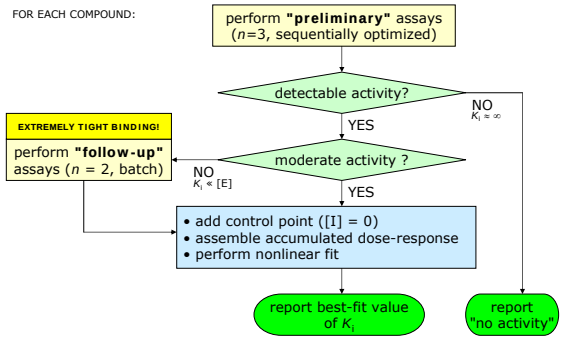
PUT TOGETHER **OPTIMAL DESIGN** AND THE **SINGLE-POINT METHOD**



Sequential optimal design: Overall outline

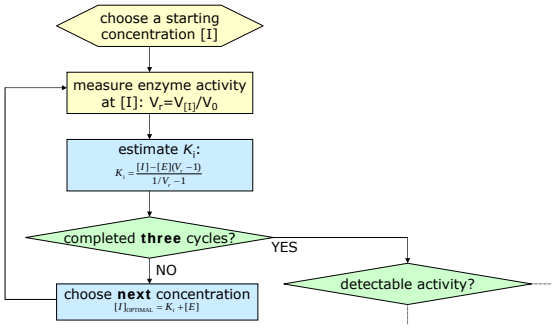
PUTTING IT ALL TOGETHER: "SINGLE-POINT METHOD" + OPTIMAL DESIGN THEORY

FOR EACH COMPOUND:



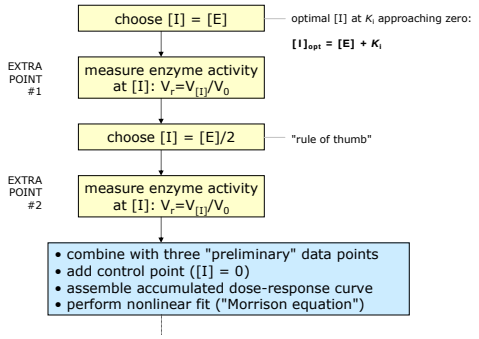
Sequential optimal design: Preliminary phase

ASSAY EVERY COMPOUND AT **THREE** DIFFERENT CONCENTRATIONS



Sequential optimal design: Follow-up phase

WE DO THIS ONLY FOR **EXTREMELY TIGHT BINDING** COMPOUNDS (K_i << $[E]_{tot}$)



Sequential optimal design: The gory details

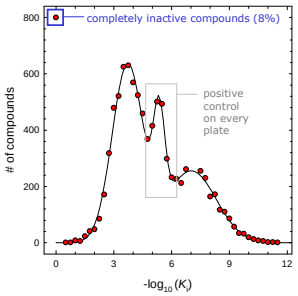
The actual "designer" algorithm is more complex:

- We need safeguards against concluding too much from *marginal data*:
  - greater than 95% inhibition, or
  - less than 5% inhibition.
- We need safeguards against falling outside the *feasible concentration range*.
- We use other safeguards and rules of thumb.
- The overall algorithm is a hybrid creation:
  - rigorous theory, and
  - practical rules, learned over many years of consulting work.

Anatomy of a screening campaign: K_i Distribution

A REAL-WORLD SCREENING PROGRAM AT **AXYS PHARMA** (LATER **CELERA GENOMICS**)

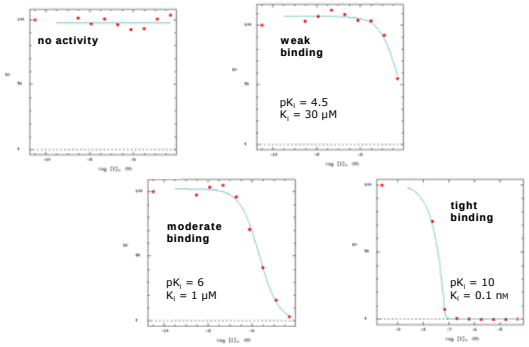
DATA COURTESY CRAIG HILL & JAMES JANC, CELERA GENOMICS
PRESENTED IN PART (BY P.K.) AT 10TH ANNUAL **SOCIETY FOR BIOMOLECULAR SCREENING**, ORLANDO, 2004



- **10,008** dose response curves
- Maximum concentration 0.5–50 µM
- Serial dilution ratio 1:4
- Eight data points per curve
- 3% Random error of initial rates
- Enzyme concentration 0.6–10 nM

Anatomy of a screening campaign: Examples

A REAL-WORLD SCREENING PROGRAM AT **AXYS PHARMA** (LATER **CELERA GENOMICS**)



Monte-Carlo simulation: Virtual sequential screen

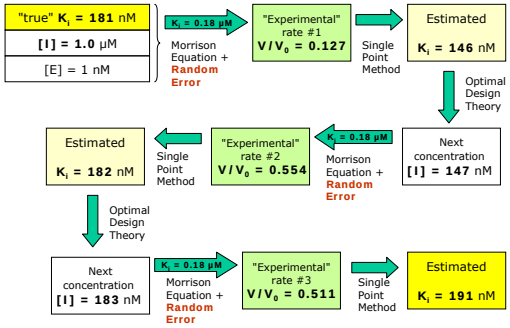
SIMULATE A POPULATION OF INHIBITORS THAT MATCHES THE AXYS/CELERA CAMPAIGN

PLAN OF A HEURISTIC *MONTE-CARLO* SIMULATION STUDY:

1. Simulate 10,000 pK_i values that match Celera's "two-Gaussian" distribution
2. Simulate enzyme activities assuming 3% random experimental error
3. Virtually "screen" the 10,000 compounds using the **sequential optimal** method
4. Compare resulting 10,000 pK_i values with the "true" (assumed) values
5. Repeat the virtual "screen" using the classic **serial dilution** method
6. Compare accuracy and efficiency of sequential and serial-dilution methods

Monte-Carlo study: Example 1 - Preliminary phase

A TYPICAL **MODERATELY POTENT** (SIMULATED) ENZYME INHIBITOR



Monte-Carlo study: Example 1 - Regression phase

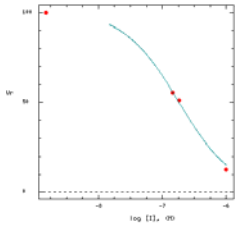
A TYPICAL **MODERATELY POTENT** (SIMULATED) ENZYME INHIBITOR - *CONTINUED*

"true" K_i = **181** nM
[E] = 1 nM
V_0 = **100**

from preliminary phase

ASSEMBLE AND FIT DOSE-RESPONSE CURVE

#	[I], µM	Rate	note
1	**0.0**	**100**	negative control
2	**1.0**	**12.7**	arbitrary initial [I]
3	**0.147**	**55.4**	optimally designed [I]
4	**0.183**	**51.1**	optimally designed [I]



from nonlinear regression

K_i = **(178 ± 9)** nM

Monte-Carlo study: Example 2 - Regression phase

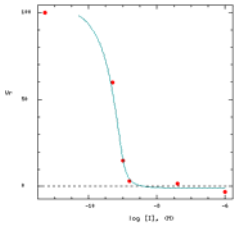
A TYPICAL **TIGHT-BINDING** (SIMULATED) ENZYME INHIBITOR

"true" K_i = **0.021** nM
[E] = 1 nM
V_0 = **100**

from preliminary phase

ASSEMBLE AND FIT DOSE-RESPONSE CURVE

#	[I], µM	Rate	note
1	**0.0**	**100**	negative control
2	**1.0**	**-3.3**	arbitrary initial [I]
3	**0.04**	**1.6**	maximum jump 25×
4	**0.0016**	**3.1**	maximum jump 25×
5	**0.001**	**13.1**	optimally designed
6	**0.0005**	**49.5**	rule of thumb



from nonlinear regression

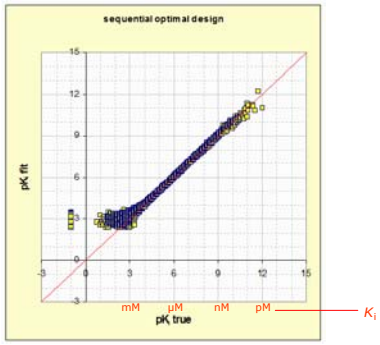
K_i = **(0.033 ± 0.011)** nM

Monte-Carlo study: "True" vs. estimated pK_i values

DISTRIBUTION OF "TRUE" pK_i VALUES IS SIMILAR TO THE AXYS/CELERA CAMPAIGN

SEQUENTIAL OPTIMAL DESIGN

n = 3(or 5) + control



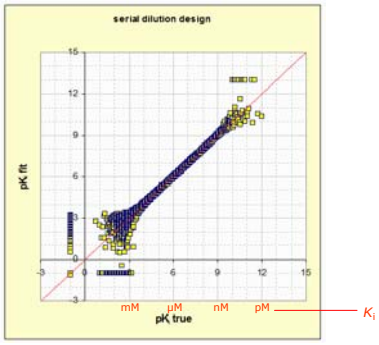
Monte-Carlo study: Dilution series results

DISTRIBUTION OF "TRUE" pK_i VALUES IS SIMILAR TO THE AXYS/CELERA CAMPAIGN

SERIAL DILUTION DESIGN

n = 8 + control

- $[I]_{max}$ = 50 µM
- Dilution 4×
- Eight wells



Efficiency of serial dilution vs. sequential design

HOW MANY WELLS / PLATES DO WE END UP USING?

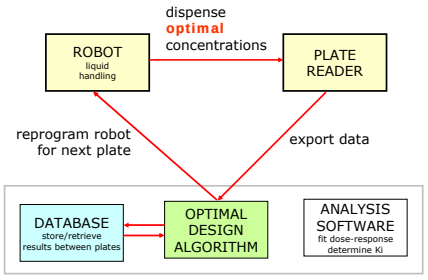
SCREEN **10,000 COMPOUNDS** (DOSE-RESPONSE) TO DETERMINE K_i's

	SERIAL DILUTION	SEQUENTIAL DESIGN	SAVINGS
total 96-well plates	909	343	62.3 %
compounds per plate	11	88	
control wells per plate	8	8	
wells with inhibitors	79992	30042	62.4 %
control wells ([I] = 0)	7272	2744	62.3 %
total wells	87264	32786	62.4 %
wells per compound	8.73	3.28	62.4 %

Toward optimized screening: Preliminary phase

PROPOSAL FOR FULLY AUTOMATED *OPTIMIZED* SCREENING

1. Accumulate minimal (optimized) dose-response curves



COMPUTER SUBSYSTEM: INSTRUMENT-CONTROL & DATA-ANALYSIS

Efficiency comparison: ~100 compounds to screen

HOW MANY WELLS / PLATES DO WE END UP USING WITH FEWER COMPOUNDS TO SCREEN?

SCREEN **88 COMPOUNDS** (DOSE-RESPONSE) TO DETERMINE K_i's

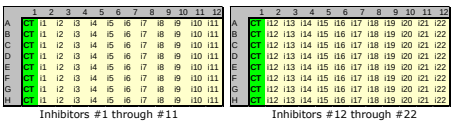
	SERIAL DILUTION	SEQUENTIAL DESIGN	SAVINGS
total 96-well plates	8	3	62.5 %
compounds per plate	11	88	
control wells per plate	8	8	
wells with inhibitors	704	264	62.5 %
control wells ([I] = 0)	64	24	62.5 %
total wells	768	288	62.5 %
wells per compound	8.73	3.27	62.5 %

Example: Plate layout for 88 inhibitors

HOW MANY WELLS / PLATES DO WE END UP USING WITH FEWER COMPOUNDS TO SCREEN?

SERIAL DILUTION

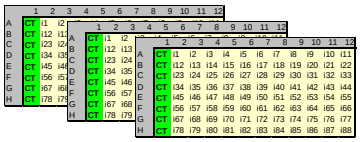
CT = control



Inhibitors #1 through #11 | Inhibitors #12 through #22 | Etc., through #88

..., 8 plates

SEQUENTIAL DESIGN



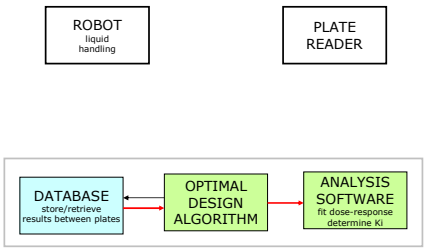
All 88 inhibitors on every plate.

3 consecutive plates with progressively optimized concentrations.

Toward optimized screening: Data-analysis phase

PROPOSAL FOR A FULLY AUTOMATED *OPTIMIZED* SCREENING

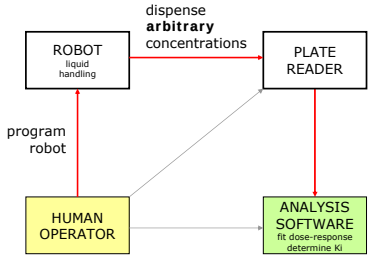
2. Analyze accumulated data



COMPUTER SUBSYSTEM: INSTRUMENT-CONTROL & DATA-ANALYSIS

Toward optimized screening: Current status

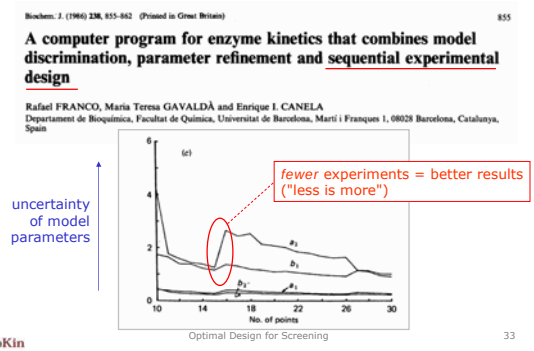
THE WAY WE SCREEN TODAY:



Optimal design in biochemistry: Earlier reports

SEARCH KEYWORDS: "OPTIMAL DESIGN", "OPTIMUM DESIGN", "OPTIM* EXPERIMENT DESIGN"

Franco *et al.* (1986) *Biochem. J.* **238**, 855-862



Biochem. J. (1986) 238, 855-862 (Printed in Great Britain) 855

A computer program for enzyme kinetics that combines model discrimination, parameter refinement and sequential experimental design

Rafael FRANCO, Maria Teresa GAVALDÀ and Enrique I. CANELA
Departament de Bioquímica, Facultat de Química, Universitat de Barcelona, Martí i Franques 1, 08028 Barcelona, Catalunya, Spain

uncertainty of model parameters

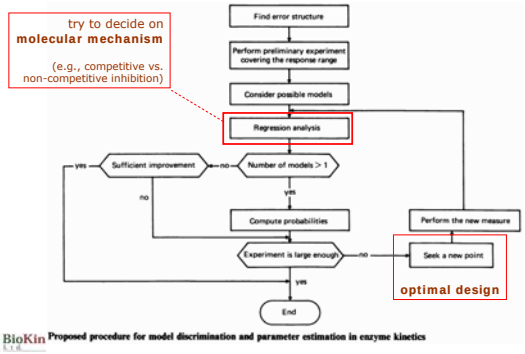
fewer experiments = better results ("less is more")

Optimal experiments for model discrimination

OPTIMAL DESIGN IS IMPORTANT FOR **MECHANISTIC ANALYSIS**

Franco *et al.* (1986) *Biochem. J.* **238**, 855-862

try to decide on **molecular mechanism** (e.g., competitive vs. non-competitive inhibition)



Proposed procedure for model discrimination and parameter estimation in enzyme kinetics

Integration with the *BatchKi* software

THE **BATCHKI** SOFTWARE IS WELL SUITED FOR PROCESSING "SMALL", OPTIMAL DATA SETS

- *Automatic* initial estimates of model parameters
  Kuzmič *et al.* (2000) *Anal. Biochem.* **281**, 62-67
- *Automatic* active-site titration (for ultra-tight binding compounds)
  Kuzmič *et al.* (2000) *Anal. Biochem.* **286**, 45-50
- *Automatic* detection of chemical impurities in samples
  Kuzmič *et al.* (2003) *Anal. Biochem.* **319**, 272-279
- *Automatic* handling of outlier data points ("Robust Regression")
  Kuzmič *et al.* (2004) *Meth. Enzymol.* **281**, 62-67

ALGORITHMS — theoretical foundation

- Handles enzyme inhibition *and* cell-based assays
- Fifteen years of experience
- Approximately 100,000 compounds analyzed by this consultant alone

Conclusions

SEQUENTIAL OPTIMAL DESIGN FOR INHIBITOR SCREENING HAS BEEN TESTED "IN SILICO"

Advantages of sequential optimal design:

- reduce material expenditures by more than 50%
- reduce screening time by more than 50%
- increase accuracy & precision of the final answer (K_i)

Disadvantages, limitations, and caveats:

- works best for large number of compounds (n > 100)
- has not been tested in practice
- to avoid programming liquid handler manually, needs "closing the loop": robot → reader → computer

Collaboration, anyone?

Acknowledgments

Craig Hill & James Janc

Theravance Inc.
South San Francisco, CA

formerly Celera Genomics – South San Francisco
formerly Axys Pharmaceuticals
formerly Arris Pharmaceuticals

Thank you for your attention

- Questions ?
- More info: www.biokin.com
- Contact: +1 617 209 4242
  +1 617 209 1616 FAX

BioKin Ltd | 15 Main Street Suite 232 | Watertown | Massachusetts 02472 | U.S.A.