Society for Biomolecular Screening
10th Annual Conference, Orlando, FL, September 11-15, 2004

Advanced Methods in

Dose-Response Screening of Enzyme Inhibitors

Petr Kuzmič, Ph.D.
BioKin, Ltd.

TOPICS:

1. **Fitting model**: Four-parameter logistic (IC_{50}) vs. Morrison equation (K_i^*)
2. **Robust regression**: Implementing outlier exclusion in practice
3. **Confidence intervals**: What should we store in activity databases?

Acknowledgements:

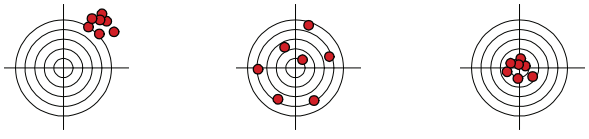
Craig Hill & Jim Janc
Celera Genomics, Department of Enzymology and HTS

BioKin Ltd.

Assumptions

- We need a *portable* measure of inhibitory potency.
- Failing portability, at least we need to *rank* compounds correctly.
- For correct ranking, we need both precision and accuracy.
- No measurement is perfectly accurate: *confidence intervals*.
- Few experiments are designed ideally and executed flawlessly.

Reminder:



PRECISION | ACCURACY | PRECISION & ACCURACY

Measures of inhibitory potency

INTRINSIC MEASURE OF POTENCY:

$$\Delta G = -RT \log K_i$$

DEPENDENCE ON EXPERIMENTAL CONDITIONS	Depends on [S]	Depends on [E]	*Example:* Competitive inhibitor
1. Inhibition constant	**NO**	**NO**	K_i
2. Apparent K_i	**YES**	**NO**	$K_i^* = K_i (1 + [S]/K_M)$
3. IC_{50}	**YES**	**YES**	$IC_{50} = K_i (1 + [S]/K_M) + [E]/2$

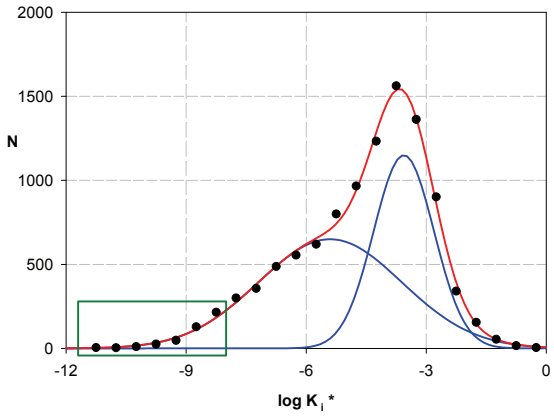
"CLASSICAL" INHIBITORS: $[E] \ll K_i$: $IC_{50} \approx K_i^*$

"TIGHT BINDING" INHIBITORS: $[E] \approx K_i$: $IC_{50} \neq K_i^*$

Tight binding inhibitors : $[E] \approx K_i$

HOW PREVALENT IS "TIGHT BINDING"?

A typical data set: ~ 10,000 **compounds**
Completely inactive: ~ 1,100 ... NOT SHOWN
Tight binding: ~ 400

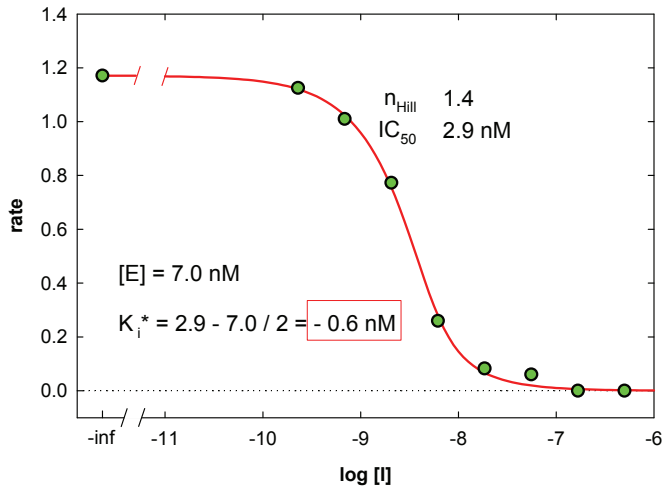


Data courtesy of Celera Genomics

Problem: Negative K_i from IC_{50}

FIT TO FOUR-PARAMETER LOGISTIC:

$K_i^* = IC_{50} - [E] / 2$

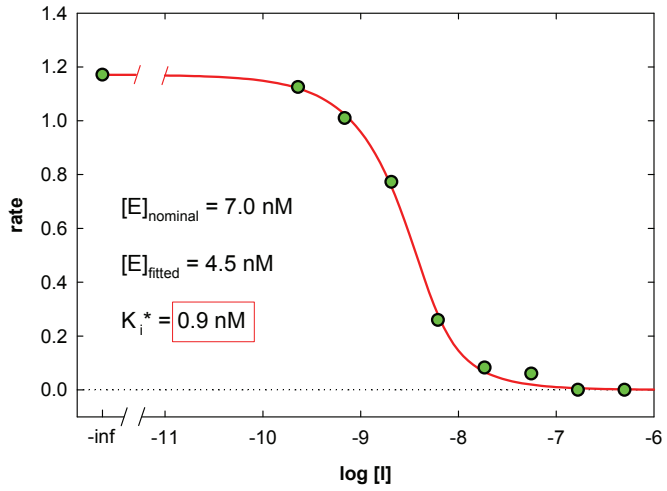


Data courtesy of Celera Genomics

Solution: Do not use four-parameter logistic

FIT TO MODIFIED MORRISON EQUATION:

P. Kuzmic *et al.* (2000) *Anal. Biochem.* 281, 62-67.
P. Kuzmic *et al.* (2000) *Anal. Biochem.* 286, 45-50.



Data courtesy of Celera Genomics

Fitting model for enzyme inhibition: Summary

MEASURE OF INHIBITORY POTENCY

- Apparent inhibition constant K_i^* is preferred over IC_{50}

MATHEMATICAL MODEL

- Modified Morrison equation is preferred over four-parameter logistic

$$v = V_b + V_0 \frac{[E]-[I]-K_i^* + \sqrt{\left([E]-[I]-K_i^*\right)^2 + 4[E]K_i^*}}{2[E]}$$

METHODOLOGY

- Optionally, adjust the enzyme concentration in fitting K_i^*

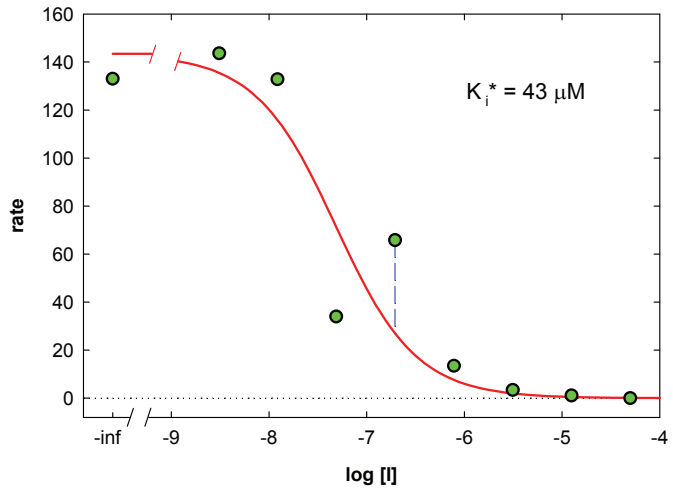
TOPICS:

1. **Fitting model**: Four-parameter logistic (IC_{50}) vs. Morrison equation (K_i^*)
2. **Robust regression**: Implementing outlier exclusion in practice
3. **Confidence intervals**: What should we store in activity databases?

Problem: Occasional "outlier" points

LEAST-SQUARES FIT

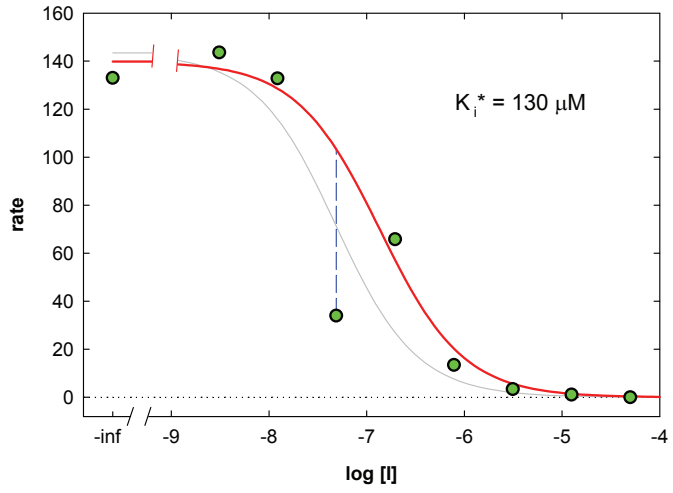
P. Kuzmic *et al.* (2004) *Meth. Enzymol.* 383, 66-81.



Solution: Robust regression ("IRLS")

HUBER'S "MINIMAX" METHOD

P. Kuzmic *et al.* (2004) *Meth. Enzymol.* 383, 66-81.



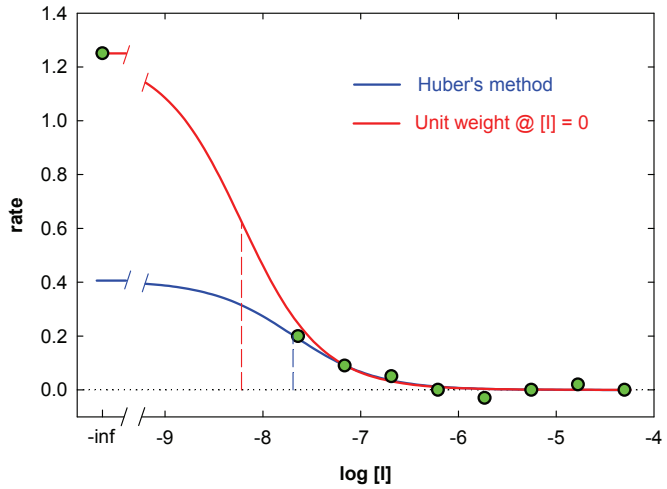
Robust fit: Practical considerations

"The devil is in the details."

- Treat negative controls in a special way (unit weight).
- Allow only a certain maximum number of "outliers".

Robust fit: Constant weighting of negative controls

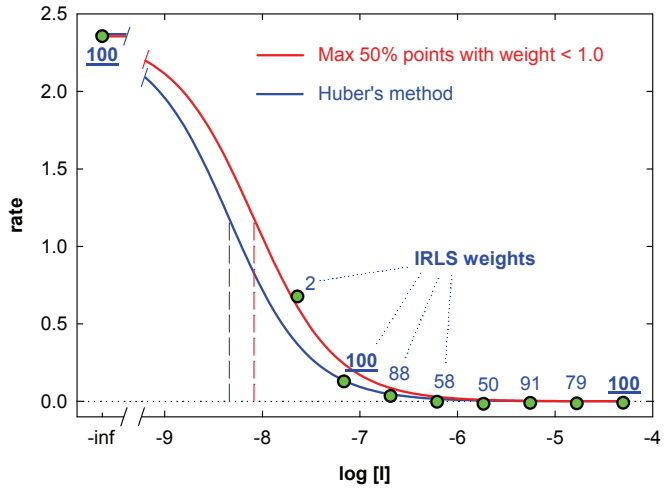
NEGATIVE CONTROL WELLS ([I] = 0) ARE EXCLUDED FROM ROBUST WEIGHTING SCHEME



Data courtesy of Celera Genomics

Robust fit: Limiting the number of "outliers"

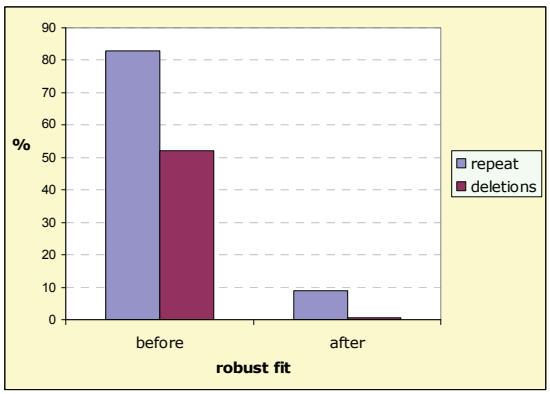
I.R.L.S.: AT MOST ONE HALF OF DATA POINTS WITH NON-UNIT WEIGHTS



Data courtesy of Celera Genomics

Robust fit: Productivity and objectivity gains

A CASE STUDY "BEFORE AND AFTER" IMPLEMENTING ROBUST REGRESSION



Data courtesy of Celera Genomics

Robust fit: Summary

- Tested on 10,000+ dose response curves
- Huber's "Minimax method" proved most effective
- Modifications for inhibitor screening:

  a. Handling of negative controls
  b. Prevent too many outliers

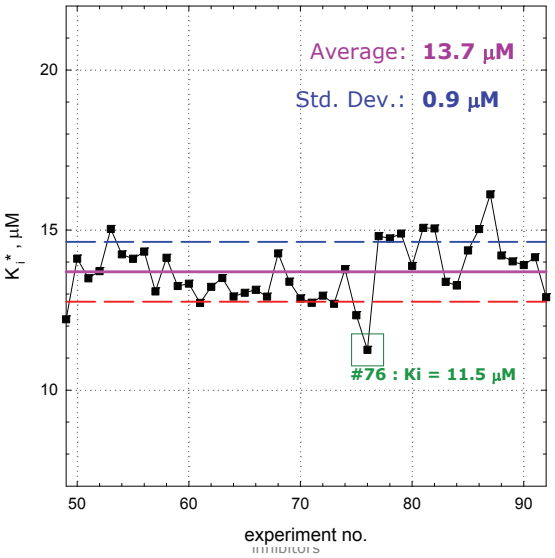
- Increase in scientific *objectivity* & *productivity*

TOPICS:

1. **Fitting model**: Four-parameter logistic (IC_{50}) vs. Morrison equation (K_i^*)
2. **Robust regression**: Implementing outlier exclusion in practice
3. **Confidence intervals**: What should we store in activity databases?

What is the "true" value of an inhibition constant?

AVERAGE & STANDARD DEVIATION FROM 43 REPLICATES



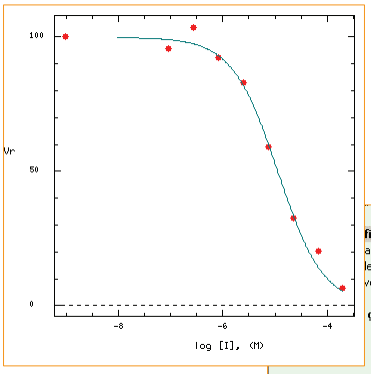
Data courtesy of Celera Genomics

BioKin Ltd.

17

Formal standard errors are too narrow

EXPERIMENT #76



Formal standard error

K_i^* = (**11.5** ± **1.2**) µM

2	Best-fit values	
3	VB	0.0
4	VO	100.3
5	E	0.0100
6	K	11.52
7	Std. Error	
8	VO	1.822
9	K	1.204
10	95% Confidence Intervals	
11	VO	95.94 to 104.6
12	K	8.674 to 14.37

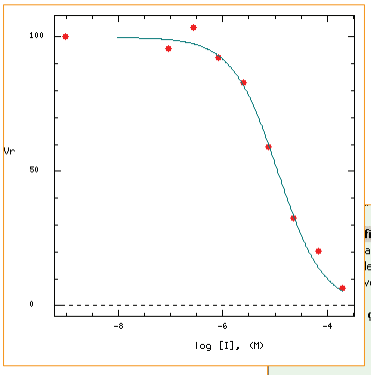
INTERVAL *DOES* ***NOT*** INCLUDE "TRUE" VALUE 13.7 µM

Data courtesy of Celera Genomics

BioKin Ltd.

Symmetrical confidence intervals are better

EXPERIMENT #76



Symmetrical 95% confidence interval

$K_i^* = (\mathbf{8.6 \ldots 14.4})\ \mu M$

2	Best-fit values	
3	VB	0.0
4	VO	100.3
5	E	0.0100
6	K	11.52
7	Std. Error	
8	VO	1.822
9	K	1.204
10	95% Confidence Intervals	
11	VO	95.94 to 104.6
12	K	8.674 to 14.37

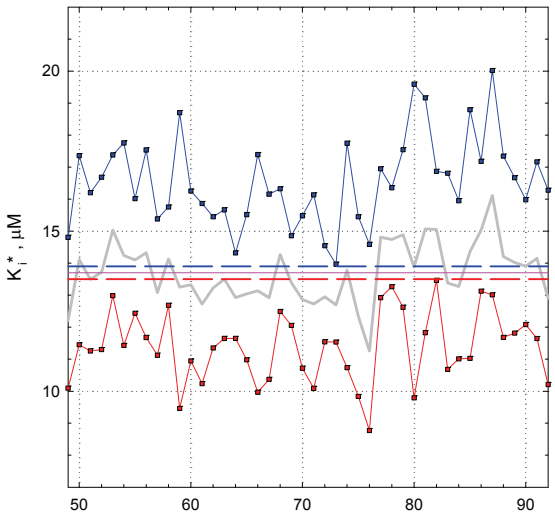
INTERVAL *DOES* INCLUDE "TRUE" VALUE 13.7 μM

Data courtesy of Celera Genomics

Nonsymmetrical confidence intervals are the best

NONSYMMETRICAL 99% C.I.

Watts, D.G. (1994) *Meth. Enzymol.* 240, 23-36.
Bates & Watts (1988) *Nonlinear Regression*, p. 207



Data courtesy of Celera Genomics

Confidence intervals (C.I.): Summary

- Report *two* numbers for each compound: *high* and *low* end of the C.I.
- If two C.I.'s overlap, the two inhibitory activities are *indistinguishable*.
- Thus, many compounds can end up with *identical rank*!

TOPICS:

1. **Fitting model**: Four-parameter logistic (IC_{50}) vs. Morrison equation (K_i*)
2. **Robust regression**: Implementing outlier exclusion in practice
3. **Confidence intervals**: What should we store in activity databases?

Conclusions: Toward a "best-practice" standard in secondary screening

BioKin Ltd.

Toward "best-practice" in secondary screening

DOSE-RESPONSE STUDIES OF ENZYME INHIBITORS

- Measure K_i^*, not IC_{50} (dependence on experimental conditions).
- Use a mechanism-based model (Morrison equation), not the four-parameter logistic equation (no physical meaning).
- Employ robust regression techniques, but very carefully.
- Report a *high/low range* (confidence interval) for every K_i^*.