environmental bioinformatics
Computational Toxicology Center

A Closed-Loop Identification Protocol (CLIP) for Nonlinear Biological Networks

Xiao-Jiang Feng, Genyuan Li, Herschel Rabitz
Department of Chemistry, Princeton University

Consortium Members

UMDNJ
ROBERT WOOD JOHNSON MEDICAL SCHOOL
University of Medicine & Dentistry of New Jersey

Computational Chemodynamics Laboratory,
Environmental & Occupational Health Sciences Institute

Department of Environmental & Occupational Medicine

Department of Pharmacology

Informatics Institute

THE STATE UNIVERSITY OF NEW JERSEY
RUTGERS

Department of Biomedical Engineering

Department of Chemical & Biochemical Engineering

Department of Environmental Sciences

Department of Statistics

Princeton University

Computer Aided Systems Laboratory,
Department of Chemical Engineering

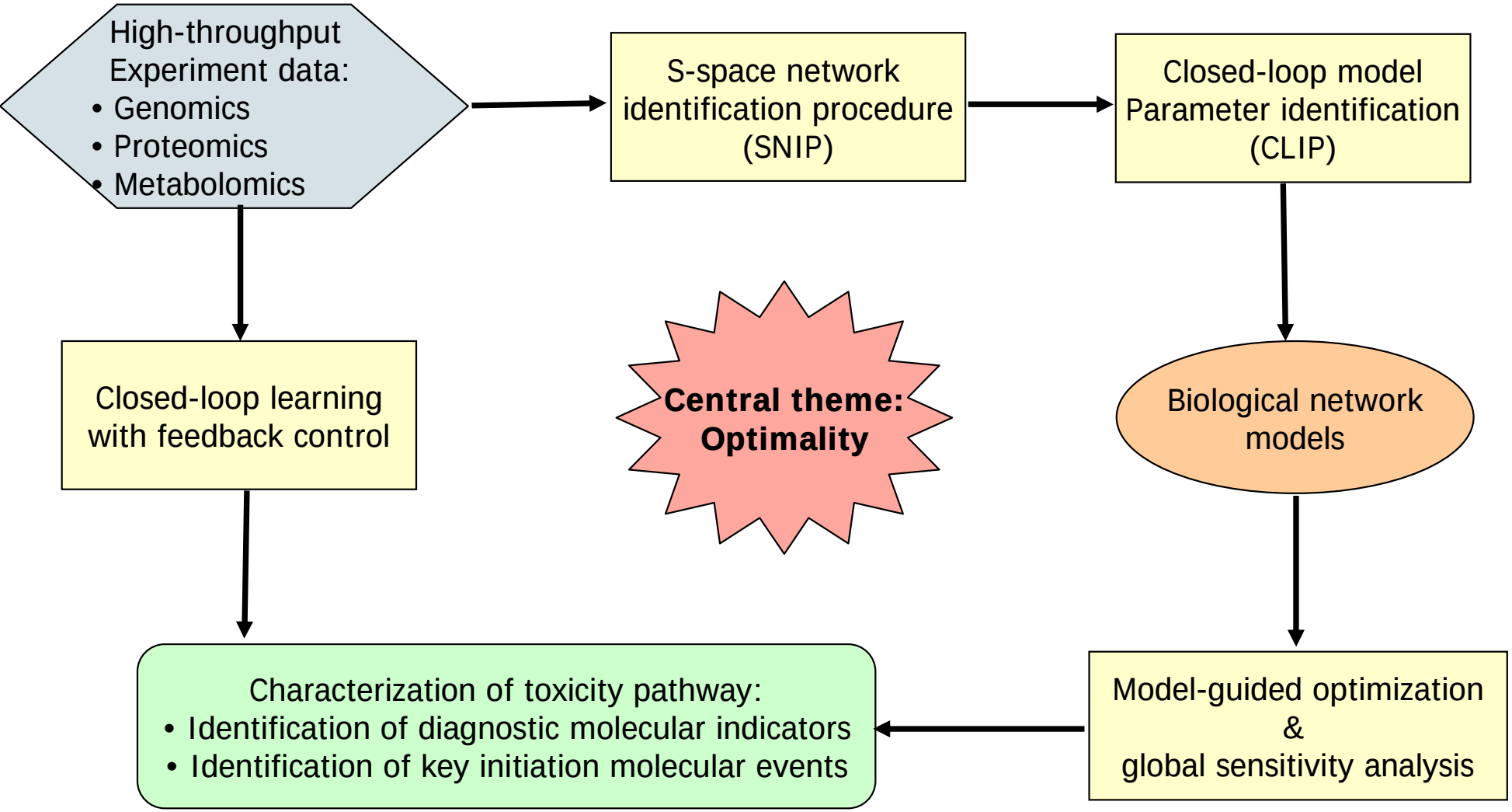
Department of Chemistry

Program in Applied and Computational Mathematics

FDA U.S. Food and Drug Administration

Center for Toxicoinformatics,
National Center for Toxicological Research

Optimal analysis, identification, and control of complex biological networks



Part A: the S-Space Network Inference Procedure (SNIP)

Objective:

- Understand who is talking to who from laboratory data
- How do they talk with each other (linear, nonlinear, independent, cooperative)?

Characteristics of the problem:

- Nonlinearity
- Dynamics mixed with network structure
- Unknown/unmeasured species
- Biological and measurement noise
- Positive and negative feedbacks, autoregulations
- Laboratory constraints
- Network scale (can be large)
- Diffusion

General strategy: perturbation + analysis

Quantitative vs. qualitative methods

Model-based vs. model-independent methods

Linear vs. nonlinear methods

Time-dependent vs. time-independent methods

Time-independent:

- **Experimentally easier**
- **Smaller information content**

Time-dependent:

- **Large information content**
- **Experimentally more difficult**
- **More difficulties in information extraction**

Existing techniques

Cluster analysis (qualitative, indirect information)

Boolean approaches (discrete assumption)

Bayesian methods (probabilistic in nature)

Dynamic modeling approaches (need model)

Correlation metric construction

- **Need random time-series perturbation**
- **Precise measurements faster than system's relaxation**
- **Perturbation around system's steady state**

Jacobian matrix methods

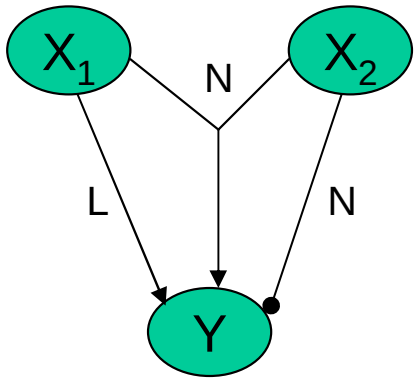
- **Linear approximation requires small perturbations, can be sensitive to noise**
- **Can become under-determined problem, need knowledge or assumptions to overcome**

The SNIP algorithm

Goals:

- Enable reliable ID from large perturbations
- Obtain linear & nonlinear interactions
- Model-independent
- Reveal independent and cooperative relationships
- Robust to noise
- Robust to unknown & unmeasured species

A toy network (1)



A three component network

Steady state relationship:

$y=x_1+3x_1x_2-5x_2^2$

Encode x_1 with $m_1(s)=\left[1+0.5\cos\left(2\times 2\pi s\right)\right]$

S is experiment index
Perturbations are time-independent

$$x_1 = x_1^* m_1(s) = 2\left[1+0.5\cos\left(2\times 2\pi s\right)\right]$$

$$y(s) = 4\cos(2\times 2\pi s)+3$$

Fourier decoding of y(s) in S

y depends linearly on x_1

A toy network (2)

Encode X_2 with $m_2(s) = [1 + 0.5\cos(5 \times 2\pi s)]$

$$y(s) = 8 + 3\cos(5 \times 2\pi s) - 5[1 + 0.5\cos(5 \times 2\pi s)]^2$$

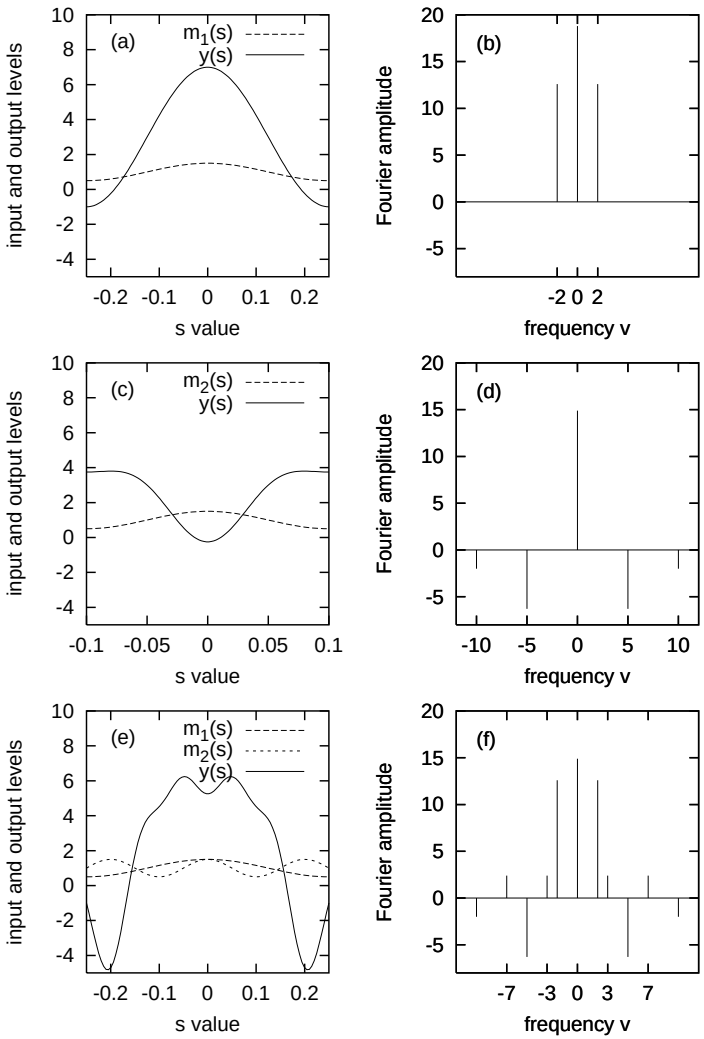
Fourier decoding of y(s) in S

y depends nonlinearly on x_2

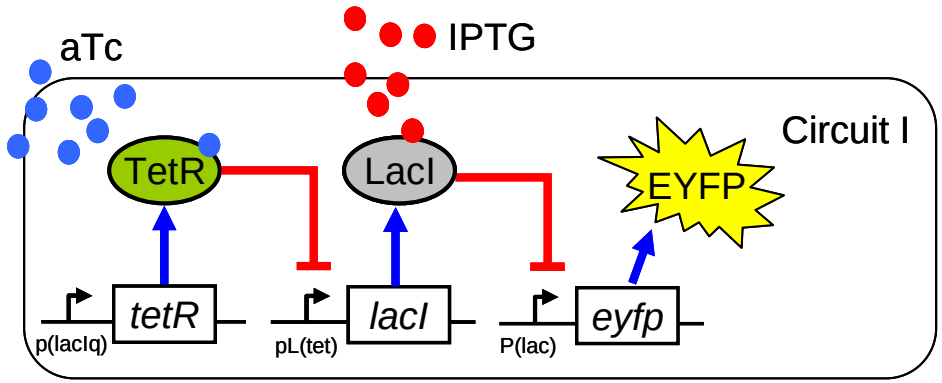
Encode X_1 (with m_1) and
X_2 (with m_2) simultaneously

Fourier decoding of y(s) in S

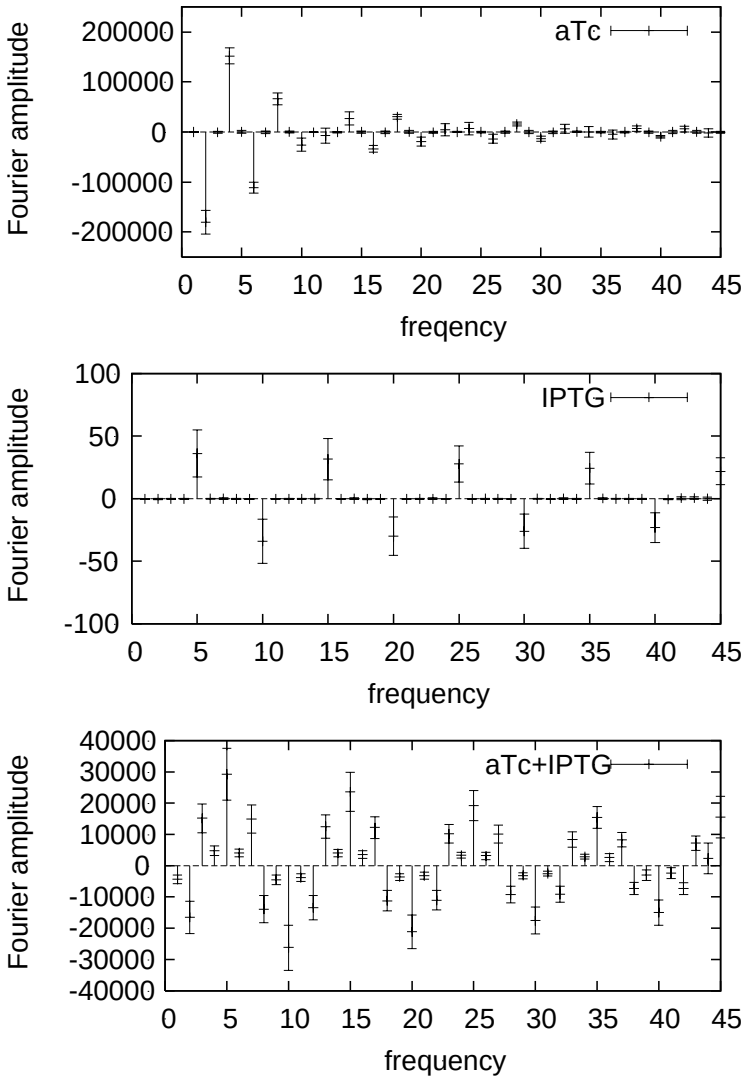
Cooperative effects of x_1 and x_2 on y



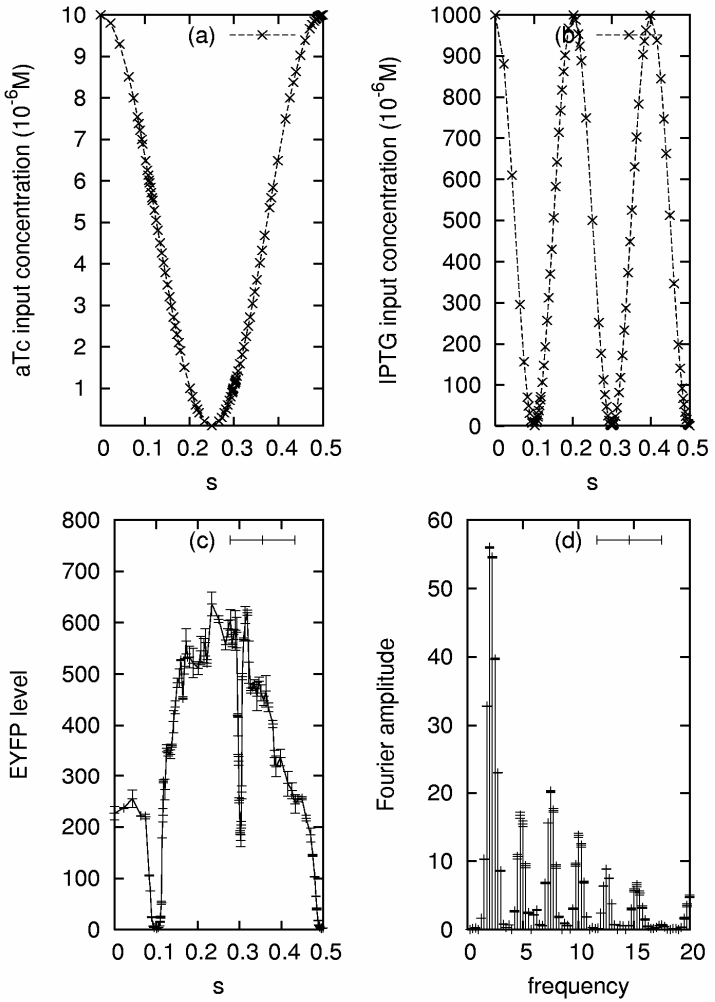
SNIP application to a simulated genetic inverter

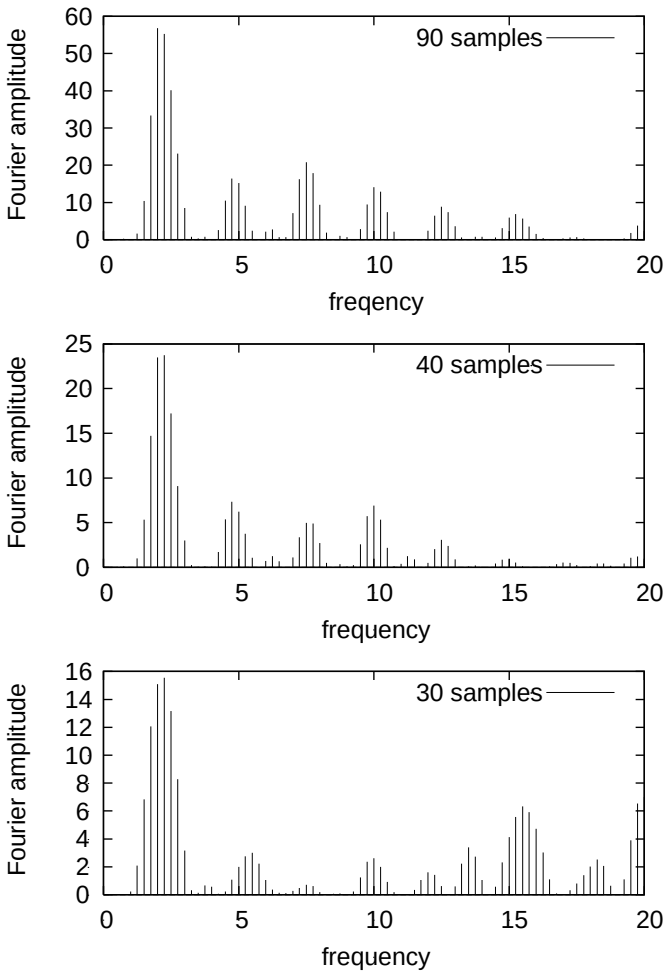


Encode: aTc & IPTG
Measure: EYFP



Laboratory SNIP application to the genetic inverter





Part B: The Closed-loop identification Protocol (CLIP)

General objective:

Optimal biochemical/biophysical model parameter identification from minimal laboratory data

Characteristics of the problem:

- System nonlinearity
- Limited number & type of experiments
- Considerable biological & measurement noise

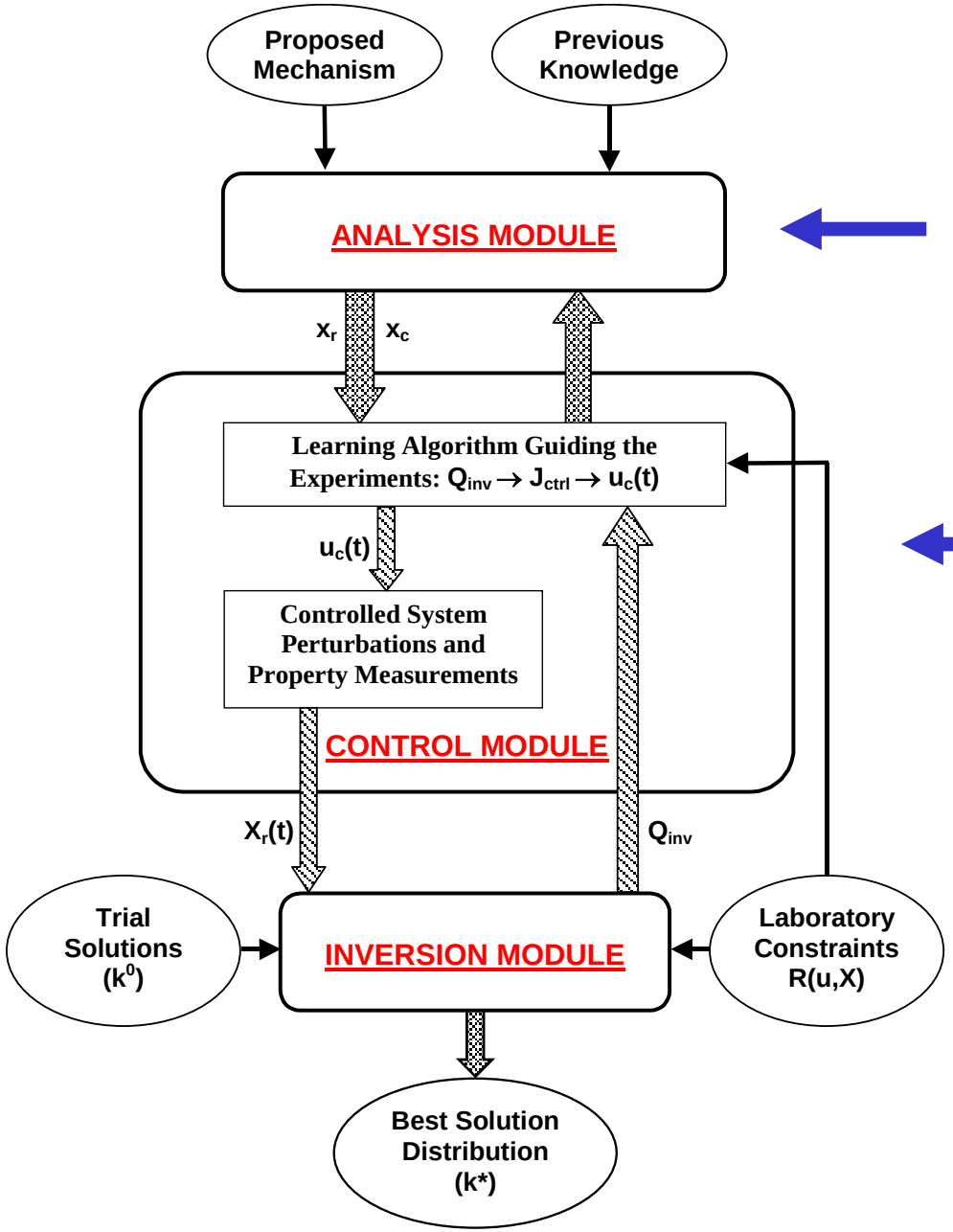
Problems with traditional identification methods:

- Provide only one or a few solutions for each parameter
- Assume linear propagation of laboratory data to inverted parameters
- Mostly based on linear system identification theory

General features of CLIP:

- Global and nonlinear identification
- Recover the full distribution of consistent solutions for each parameter
- Iteratively search for the most informative experiments
- Maximally reduce parameter uncertainty

The CLIP operation



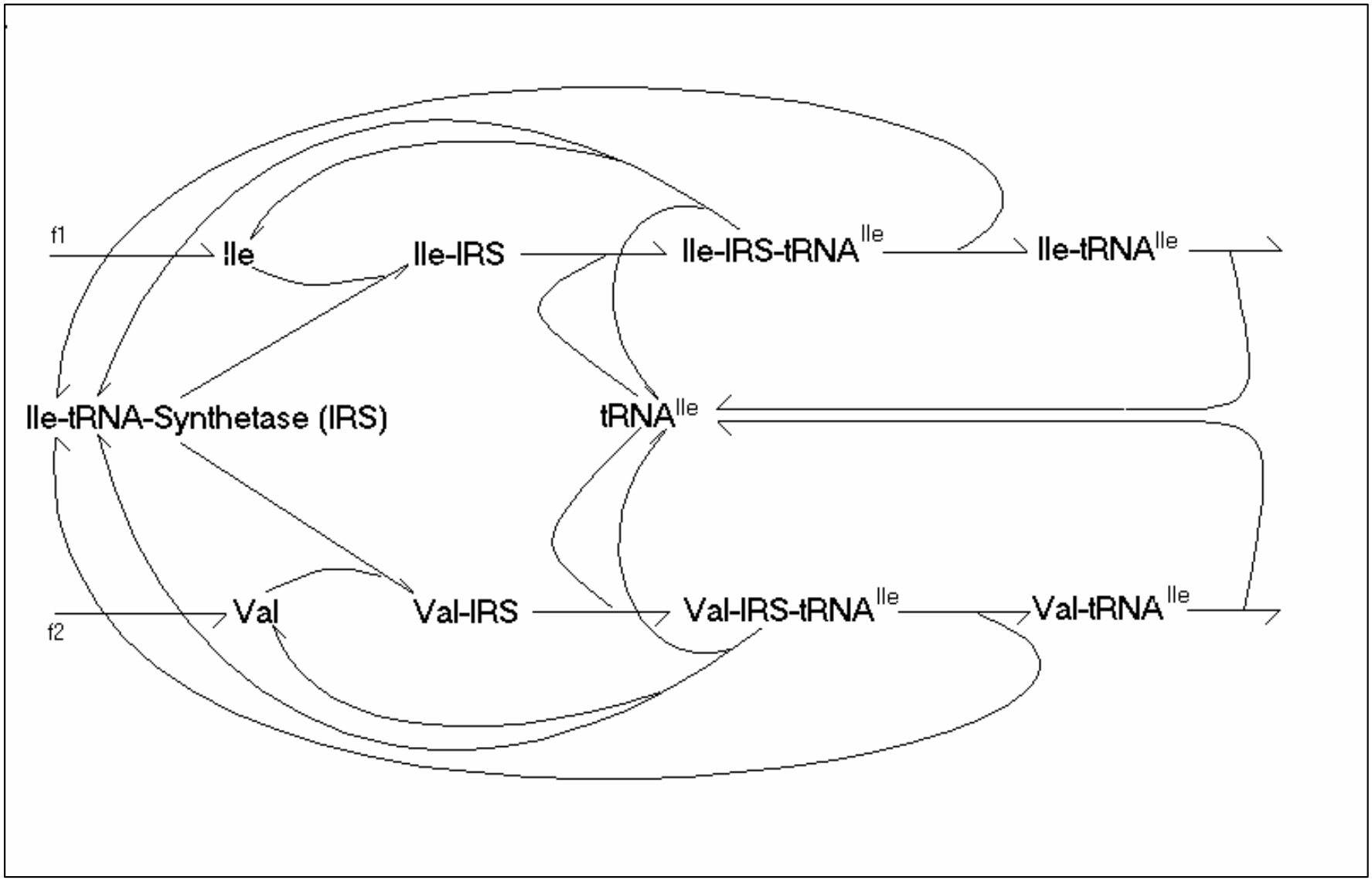
Pre-lab analysis and experimental design

Iterative experiment optimization and data acquisition

Global parameter identification

Isoleucyl-tRNA synthetase proofreading valyl-tRNA^{Ile}

M. Okamoto & M. Savageau. *Biochemistry* 23:1701-1709 (1984)



Kinetic equations (10 species, 16 rate constants)

$$dx_1/dt = k_{-3}x_5 + k_{-4}x_6 + (k_7 + k_1)x_7 + (k_8 + k_2)x_8 - k3x_1x_3 - k4x_1x_4 - k_{-7}x_1x_9 - k_{-8}x_1x_{10}$$
$$dx_2/dt = (k_{-5} + k_1)x_7 + (k_{-6} + k_2)x_8 + k_9x_9 + k_{10}x_{10} - k_5x_2x_5 - k_6x_2x_6$$
$$dx_3/dt = f_1 + k_{-3}x_5 + k_1x_7 - k_3x_1x_3$$
$$dx_4/dt = f_2 + k_{-4}x_6 + k_2x_8 - k_4x_1x_4$$
$$dx_5/dt = k_3x_1x_3 + k_{-5}x_7 - k_{-3}x_5 - k_5x_2x_5$$
$$dx_6/dt = k_4x_1x_4 + k_{-6}x_8 - k_{-4}x_6 - k_6x_2x_6$$
$$dx_7/dt = k_5x_2x_5 + k_{-7}x_1x_9 - (k_{-5} + k_7 + k_1)x_7$$
$$dx_8/dt = k_6x_2x_6 + k_{-8}x_1x_{10} - (k_{-6} + k_8 + k_2)x_8$$
$$dx_9/dt = k_7x_7 - k_{-7}x_1x_9 - k_9x_9$$
$$dx_{10}/dt = k_8x_8 - k_{-8}x_1x_{10} - k_{10}x_{10}$$

x_1 = [IRS] x_2 = [$tRNA^{Ile}$] x_3 = [Ile] x_4 = [Val] x_5 = [Ile-IRS] x_6 = [Val-IRS]
x_7 = [Ile-IRS-$tRNA^{Ile}$] x_8 = [Val-IRS-$tRNA^{Ile}$] x_9 = [Ile-$tRNA^{Ile}$] x_{10} = [Val-$tRNA^{Ile}$]

The task:

Obtain the rate constant k_1, k_2, k_5, k_{-5}, k_6, k_{-6}

Experimental capabilities and restrictions:

- Positive influx of x_1, x_3, x_4
- Finite concentration measurements
- Laboratory noise

The analysis module: estimating the most informative experiments

- **Estimate the best molecules for monitoring system behavior**
- **Determine the best molecular targets for perturbing the system**

Sensitivity analysis by Random-Sampling High Dimensional Model Representation (RS-HDMR)

$$\sigma^2_{total} = \sum_{i=1}^{n} \sigma_i^2 + \sum_{1 \le i < j \le n} \sigma_{ij}^2 + \dots$$

The inversion module: identifying the rate constant distribution

The Genetic Algorithm (GA)

Mutation

1101 1111+1100 0010

1101 1101+1100 0110

Crossover

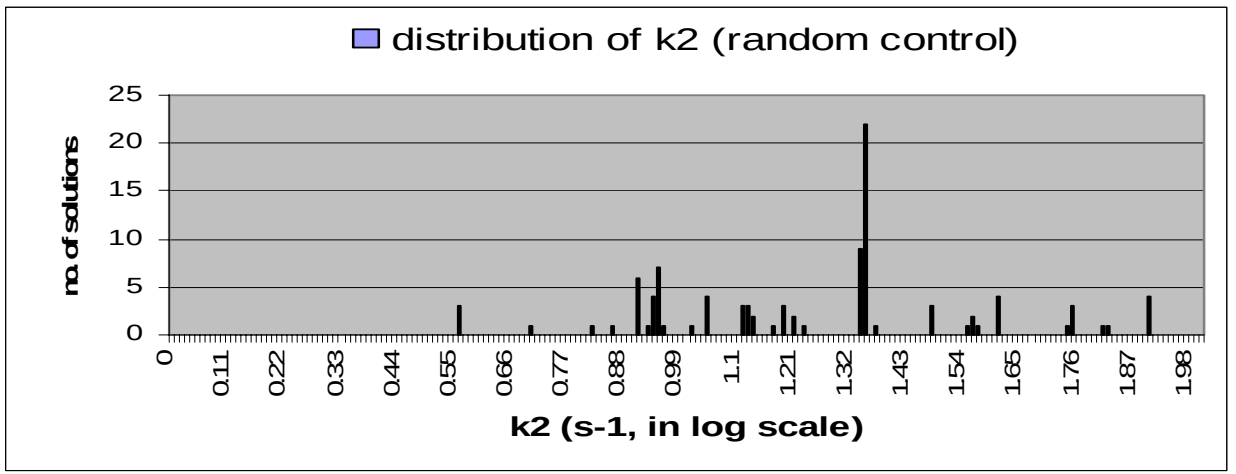
1101 1100 + 1111 **0010**

1101 **0010** + 1111 1100

Cost function of the inversion GA

$$J_{inv}^{i,p} = \frac{1}{N}\sum_{n=1}^{N} \frac{1}{T}\sum_{t=t_1}^{t_T} \left(X_{n,t}^{i,lab} - X_{n,t}^{i,p,cal}\right) / \varepsilon_n^i$$

Typical rate constant distribution after random perturbation



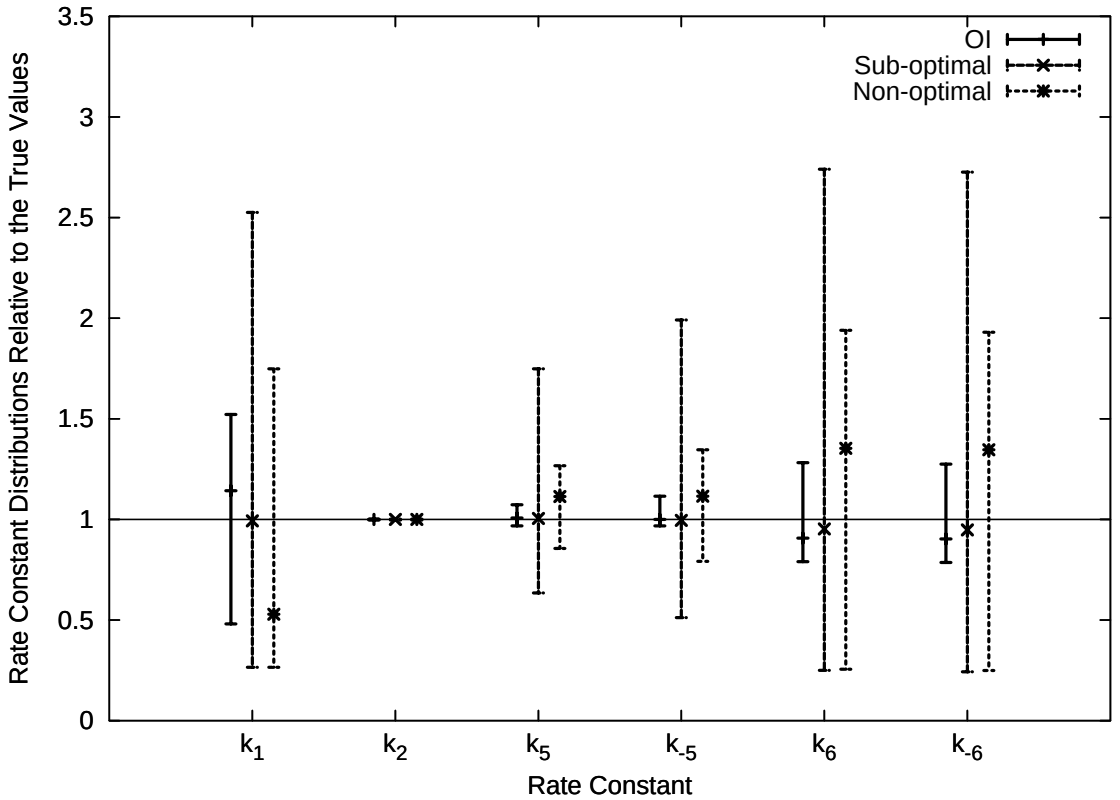
The control module: squeezing on the rate constant distribution

Cost function of the control GA

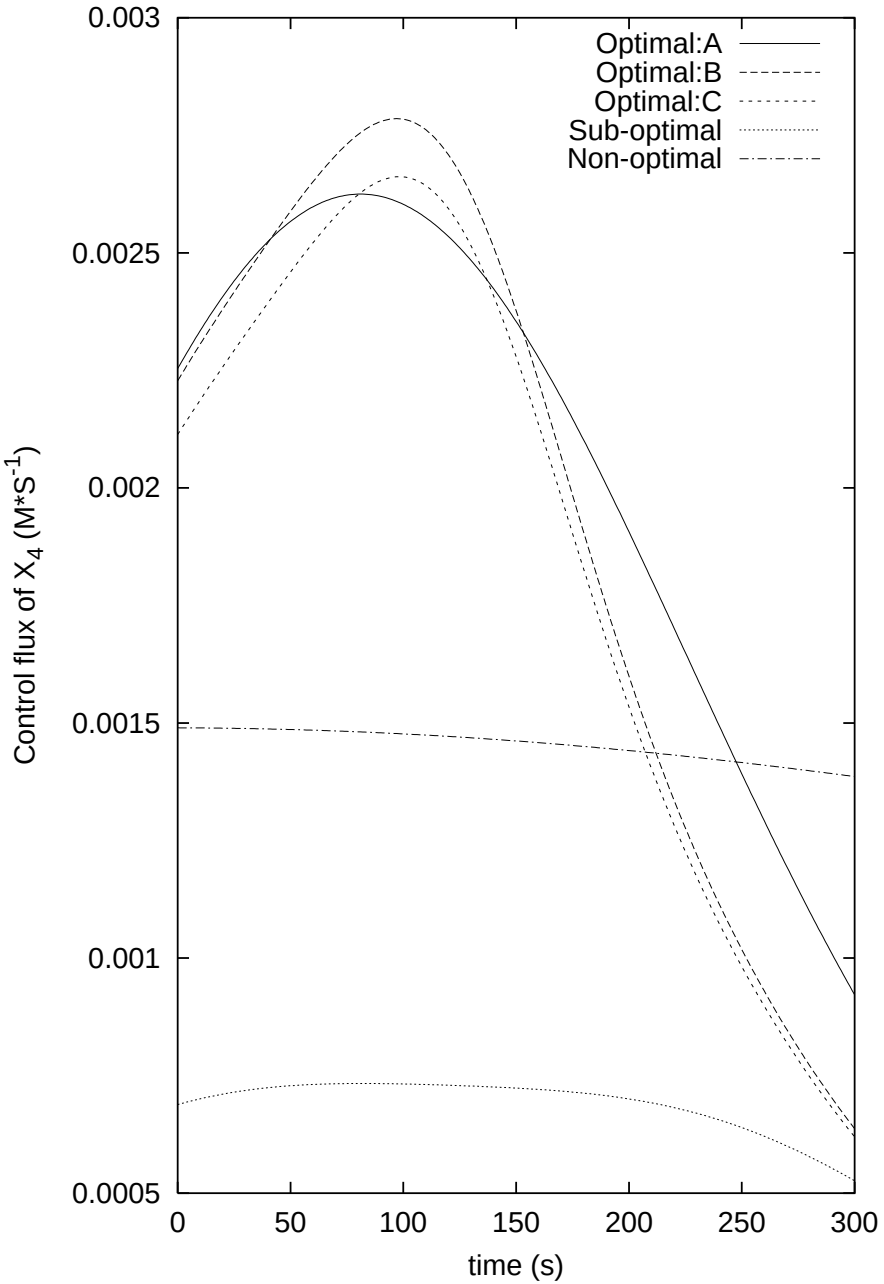
$$J^i_{ctrl} = Q^i_{inv} - \omega R[u^i_c(t), X^i_r(t)]$$

$$Q^i_{inv} = 1/[\frac{1}{M}\sum_{m=1}^{M}\frac{(k^i_{m,\max} - k^i_{m,\min})}{(k^i_{m,\max} + k^i_{m,\min})}]$$

Rate constant distribution after 500 experiments



Convergence of the optimal perturbations



Problems with CLIP:

1. Large number of experiments
2. Expensive computation

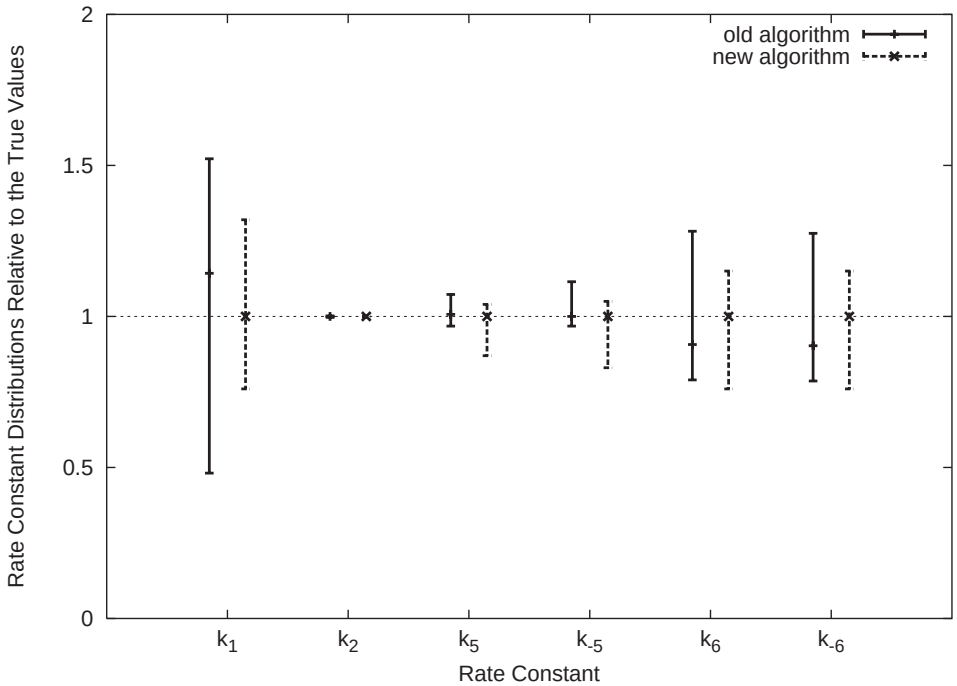
Proposed solution:

1. Replace the GA by the simplex algorithm in the control module
2. Modify the cost function in the inversion module

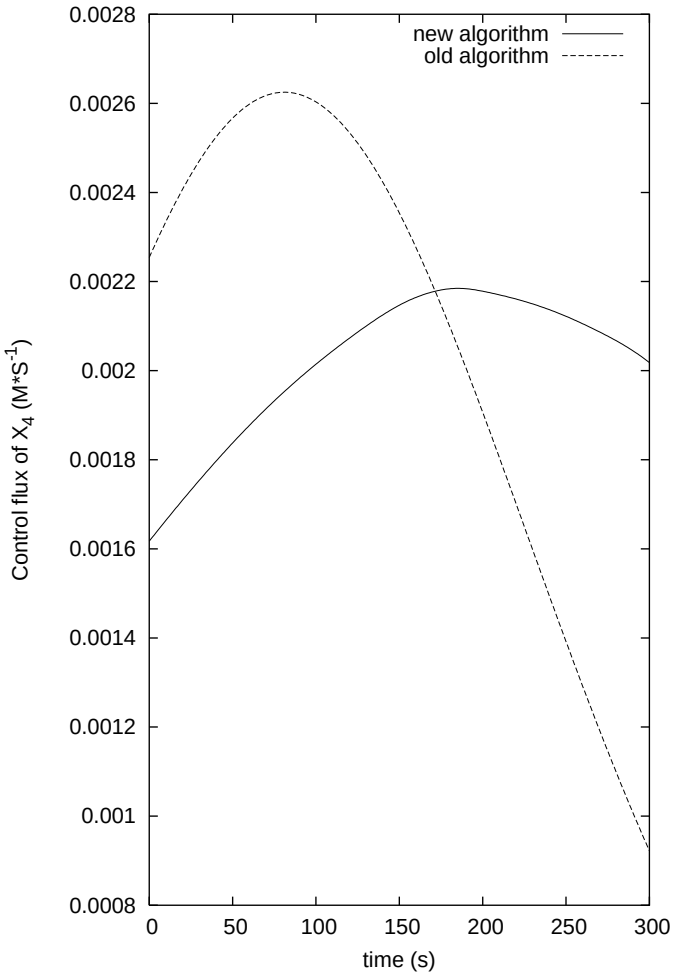
$$J_{inv}^{i,p} = \frac{1}{T}\frac{1}{N}\sum_{t=t_1}^{t_T}\sum_{n=1}^{N}\exp[(X_{n,t}^{i,lab} - X_{n,t}^{i,p,cal})/\varepsilon_n^i]\times[\prod_{m=1}^{M}(\frac{k_m^U - k_m^S + 1}{k_m^U - k_m^L + 1})^w]$$

The Simplex-CLIP algorithm

The inversion quality



Optimal perturbation



The Simplex-CLIP: 10 times less experiments, 20 times less computational cost

Problems with the simplex-CLIP:

- Simplex algorithm lacks sufficient global search ability
- Considerable number of experiments

Question: what makes an optimal perturbation optimal?

One answer: it maximizes the system's sensitivity with respect to variations in the rate constants?

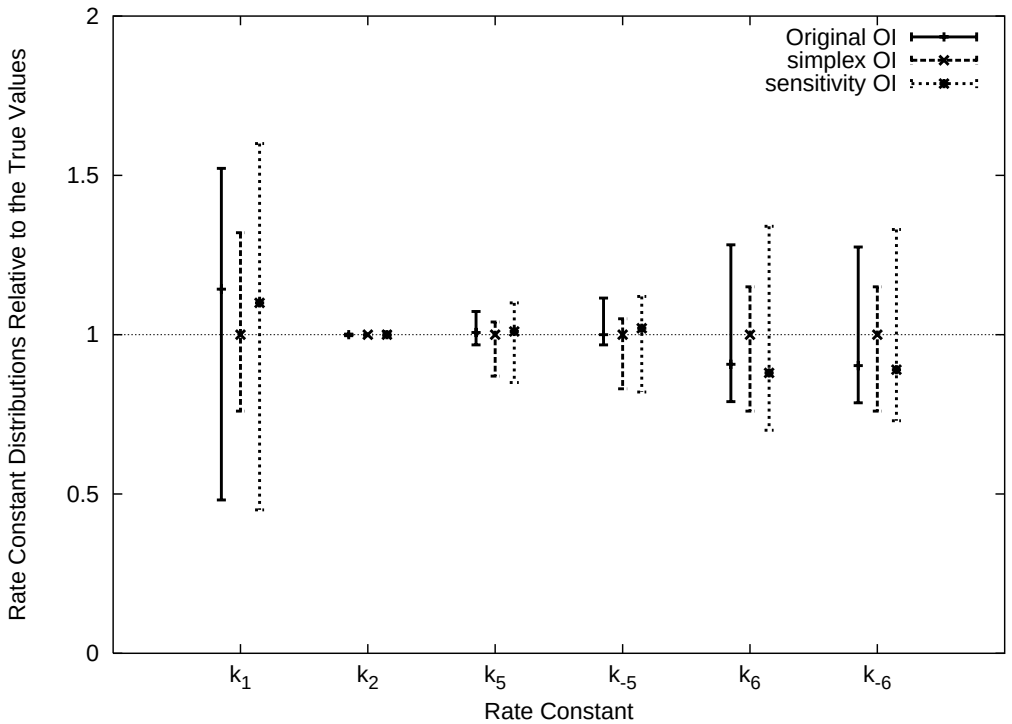
Proposed solution: estimate the optimal perturbation using global sensitivity maximization

Advantage:

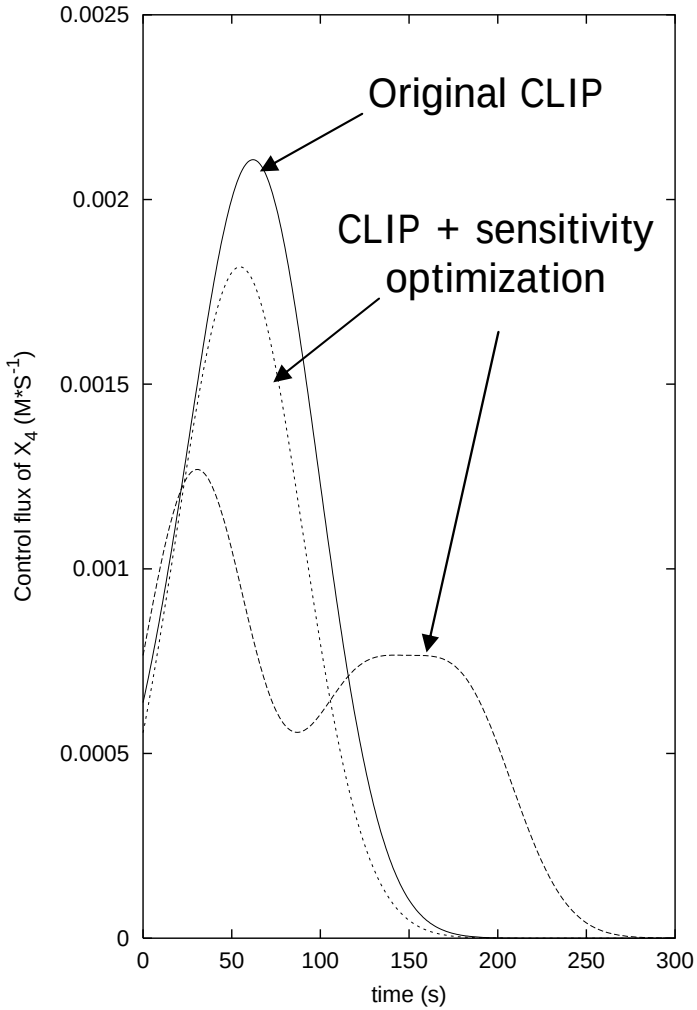
- No experiment needed for sensitivity maximization!
- Global search
- Inexpensive computation

Effect of sensitivity optimization for CLIP

Inversion quality



Optimal perturbation



CLIP + sensitivity optimization: ONE laboratory experiment and >100 times less computational time

X.Feng, H. Rabitz, et al. A closed-loop identification protocol for nonlinear dynamical systems. *J. Phys. Chem. A* **110:**7755-7762 (2006)

Future Directions

- SNIP
  - quantitative relationship extraction
  - Encoding experiment optimization
  - Simulation on larger networks (scalability)
- CLIP
  - Simulation on larger networks (scalability)
  - Model discrimination
- Application to toxicity pathways

Acknowledgements

Prof. Herschel Rabitz

Dr. Genyuan Li

Rabitz group, Princeton University

Prof. Ron Weiss, Electrical Engineering, Princeton University

Prof. Gabriel Turinici, INRIA Rocquencourt, France

Support for this work has been provided primarily by the USEPA-funded Environmental Bioinformatics and Computational Toxicology Center (ebCTC), under STAR Grant number GAD R 832721-010. This work has not been reviewed by and does not represent the opinions of the funding agency.