Data-driven parameter inference for gene circuit modeling

Linh Huynh* and Ilias Tagkopoulos
Department of Computer Science
UC Davis Genome Center
University of California, Davis
*Email: huynh@ucdavis.edu

Introduction

- **Our goal:** Accurate **model-driven** designs for synthetic biology.

UC DAVIS
UNIVERSITY OF CALIFORNIA

Introduction

- **Our goal:** Accurate **model-driven** designs for synthetic biology.
- **Challenge**
  - Lack of **accurate** parameter values

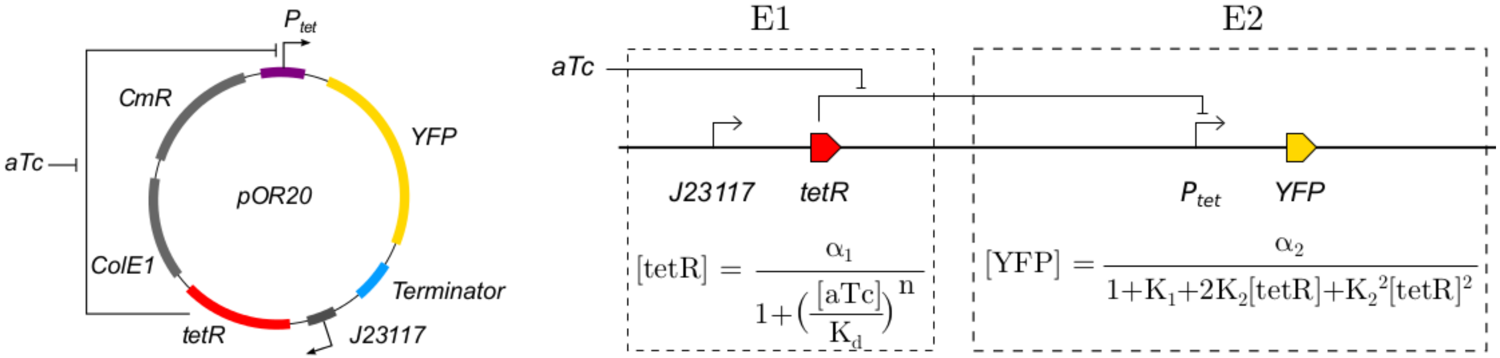
Introduction

- **Our goal:** Accurate **model-driven** designs for synthetic biology.
- **Challenge**
  - Lack of **accurate** parameter values
    - Measuring **directly** from experiment is **difficult**

Introduction

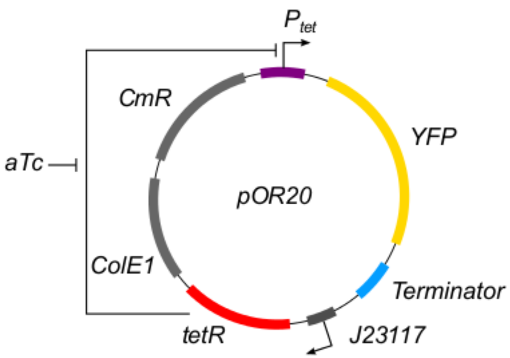
- **Our goal:** Accurate **model-driven** designs for synthetic biology.
- **Challenge**
  - Lack of **accurate** parameter values
    - Measuring **directly** from experiment is **difficult**
    - Estimating **indirectly** from experimental data is **inaccurate**

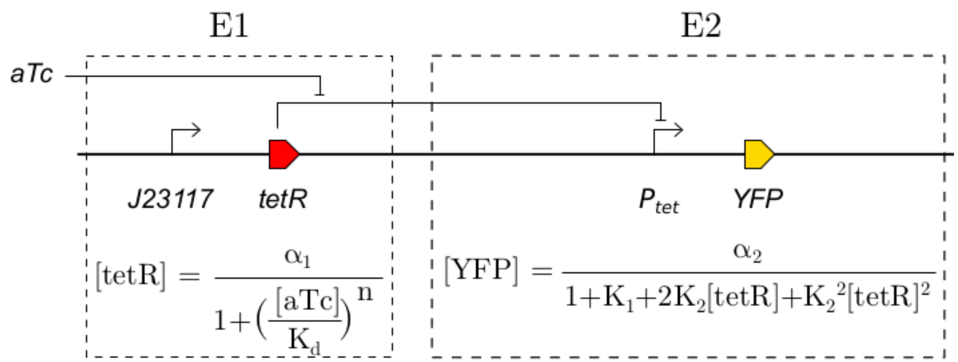
Example: Inference



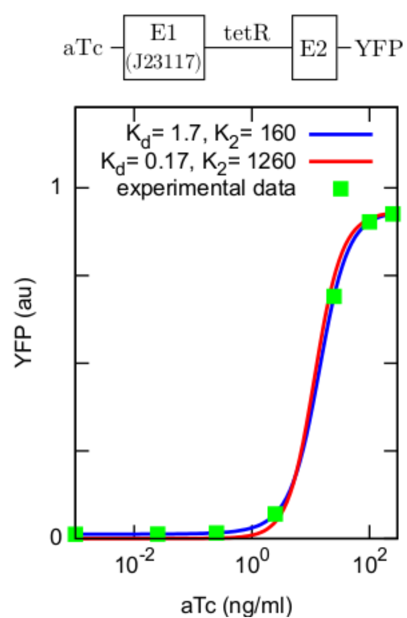
(***Tamsir et al, Nature, 2011***)

Example: Inference

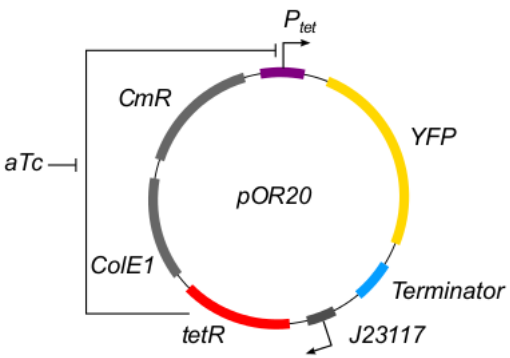


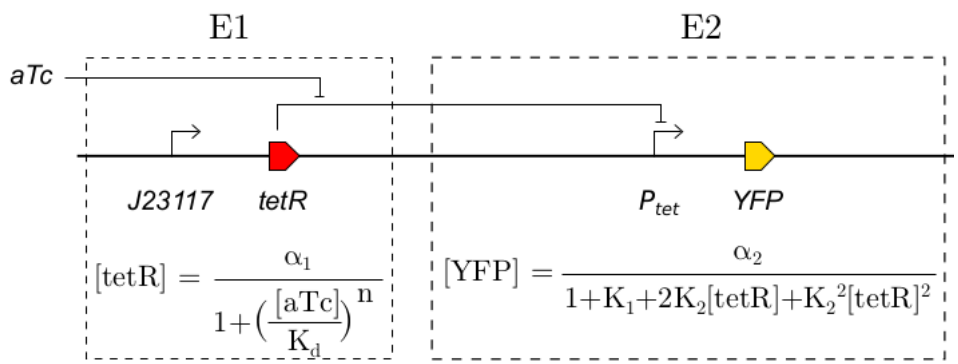


(***Tamsir et al, Nature, 2011***)

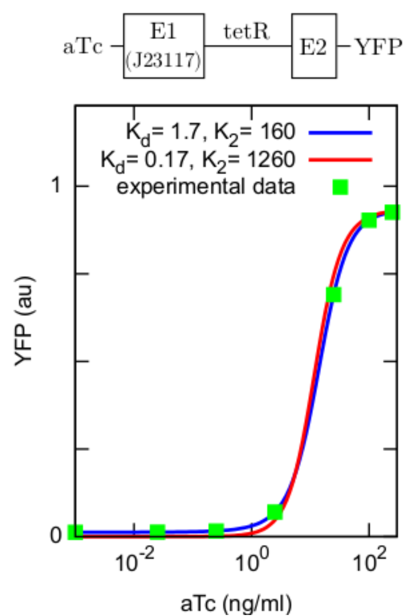


Example: Inference



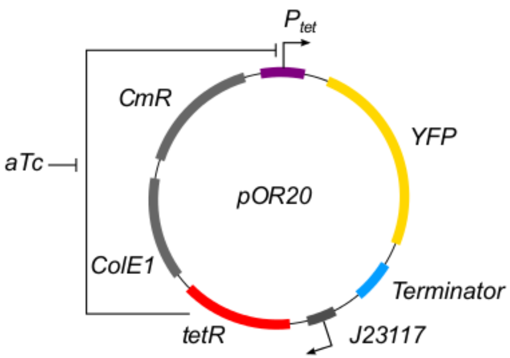


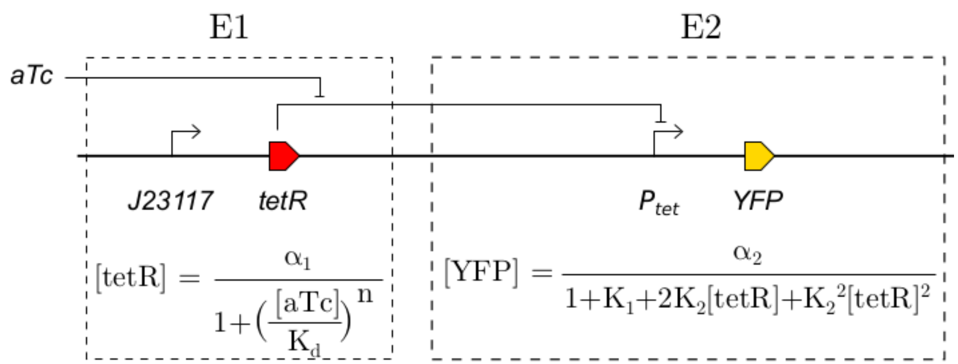
(*Tamsir et al, Nature, 2011*)



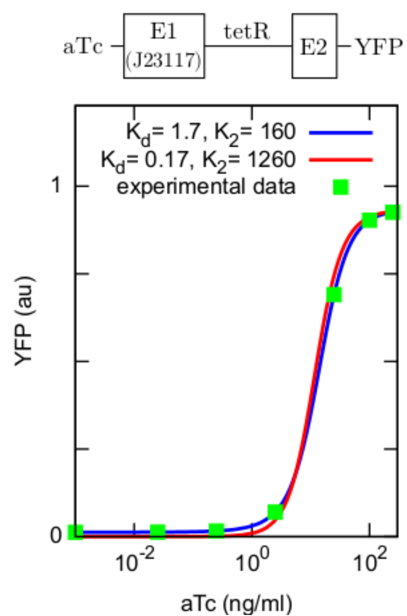
Multiple solutions
(K_d = 1.7, K_2 = 160)
vs (K_d = 0.17, K_2 = 1260)

Example: Inference





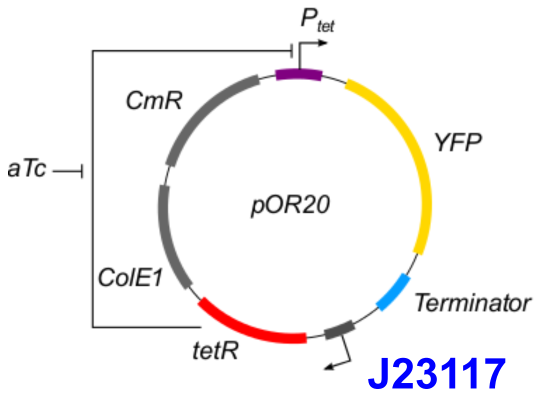
(***Tamsir et al, Nature, 2011***)



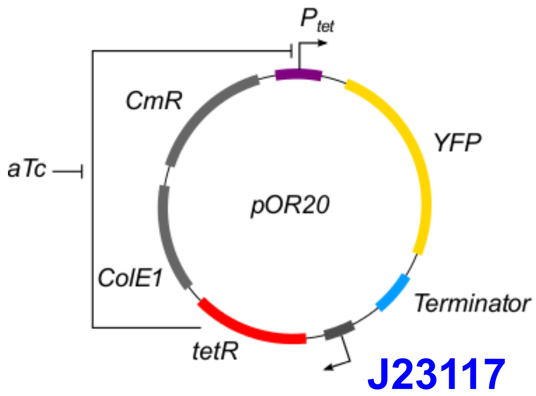
Multiple solutions
(K_d = 1.7, K_2 = 160)
vs (K_d = 0.17, K_2 = 1260)

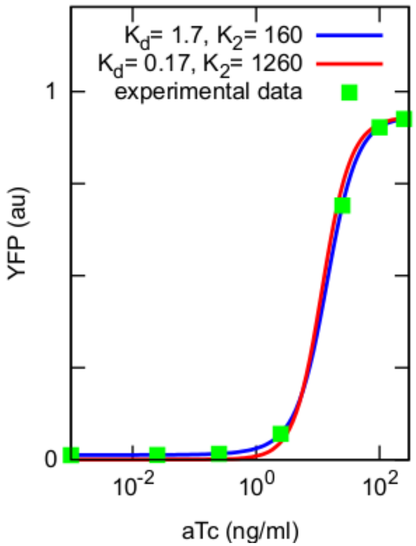
Parameters K_d, K_2 are uncertain!

Example: Design with uncertain parameters

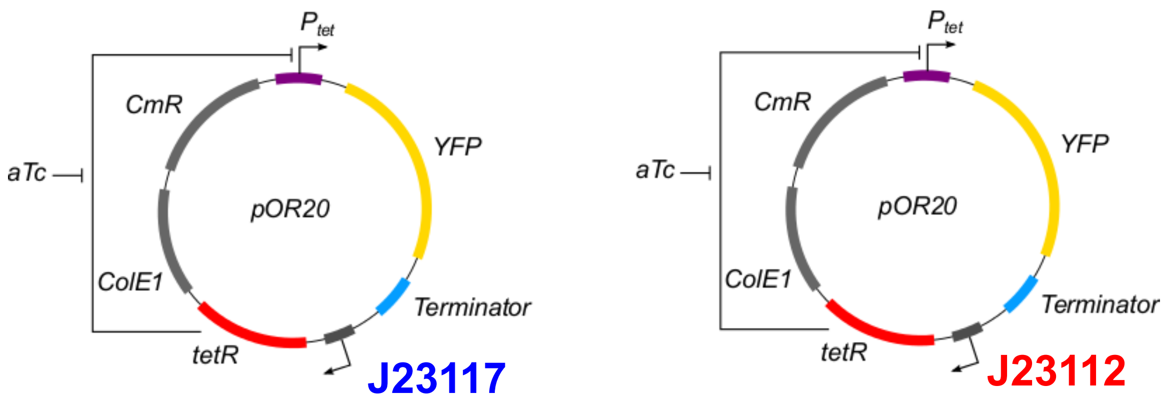


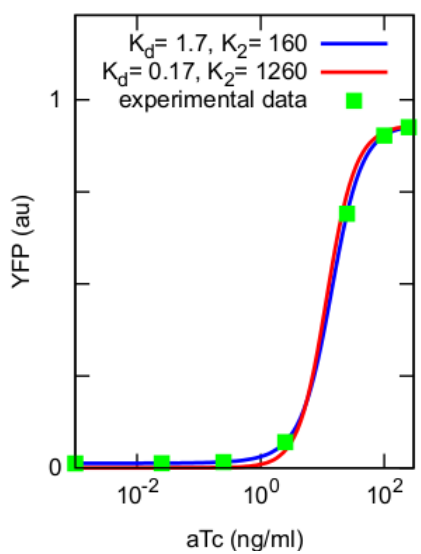
Example: Design with uncertain parameters





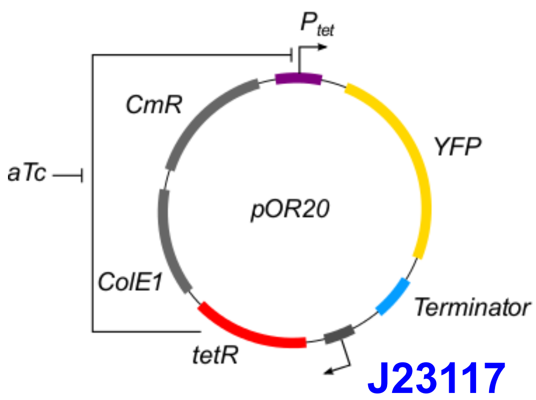
Example: Design with uncertain parameters

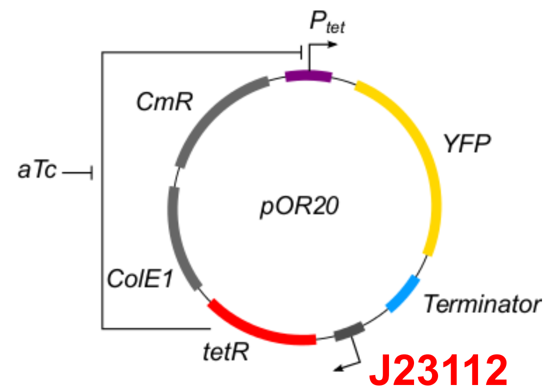


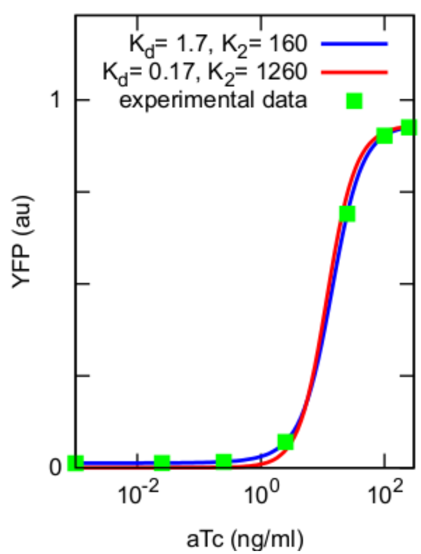


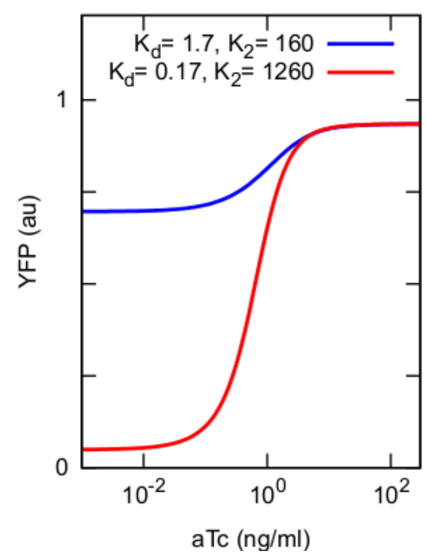
UCDAVIS
UNIVERSITY OF CALIFORNIA

Example: Design with uncertain parameters

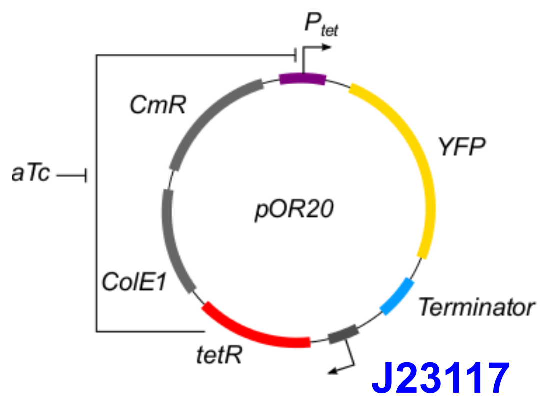


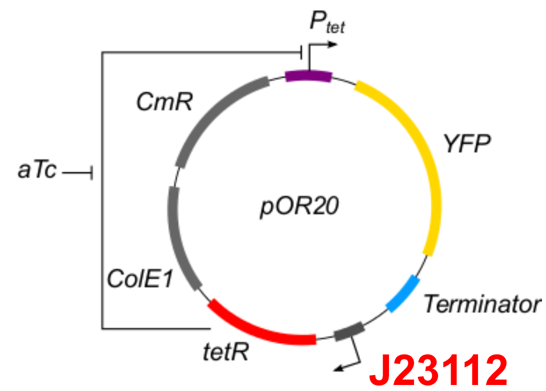


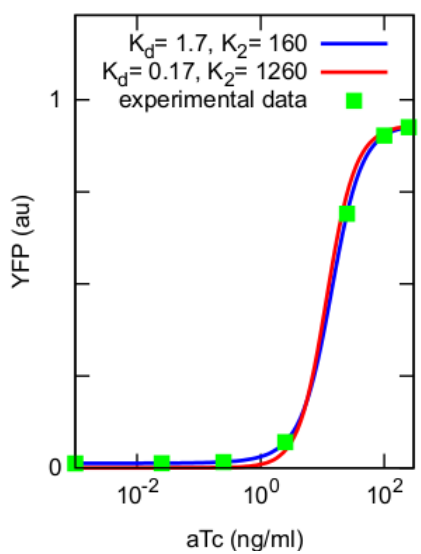


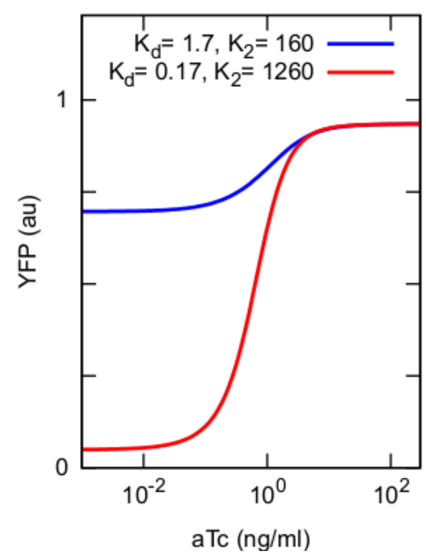


Example: Design with uncertain parameters









Circuit behavior prediction is unreliable!

Why do we have multiple fitting solutions?

Why do we have multiple fitting solutions?

- Parameters are not well-constrained

Why do we have multiple fitting solutions?

- Parameters are not well-constrained
  - The model output change due to the change of one parameter can be compensated by the change from changing other parameters.

Why do we have multiple fitting solutions?

- Parameters are not well-constrained
  - The model output change due to the change of one parameter can be compensated by the change from changing other parameters.
- Can we add more constraints for parameters?

Why do we have multiple fitting solutions?

- Parameters are not well-constrained
  - The model output change due to the change of one parameter can be compensated by the change from changing other parameters.
- Can we add more constraints for parameters?
  - Do more experiments (e.g. measure tetR)

UCDAVIS
UNIVERSITY OF CALIFORNIA

Why do we have multiple fitting solutions?

- Parameters are not well-constrained
  - The model output change due to the change of one parameter can be compensated by the change from changing other parameters.
- Can we add more constraints for parameters?
  - Do more experiments (e.g. measure tetR)
    - **Prohibitively costly**

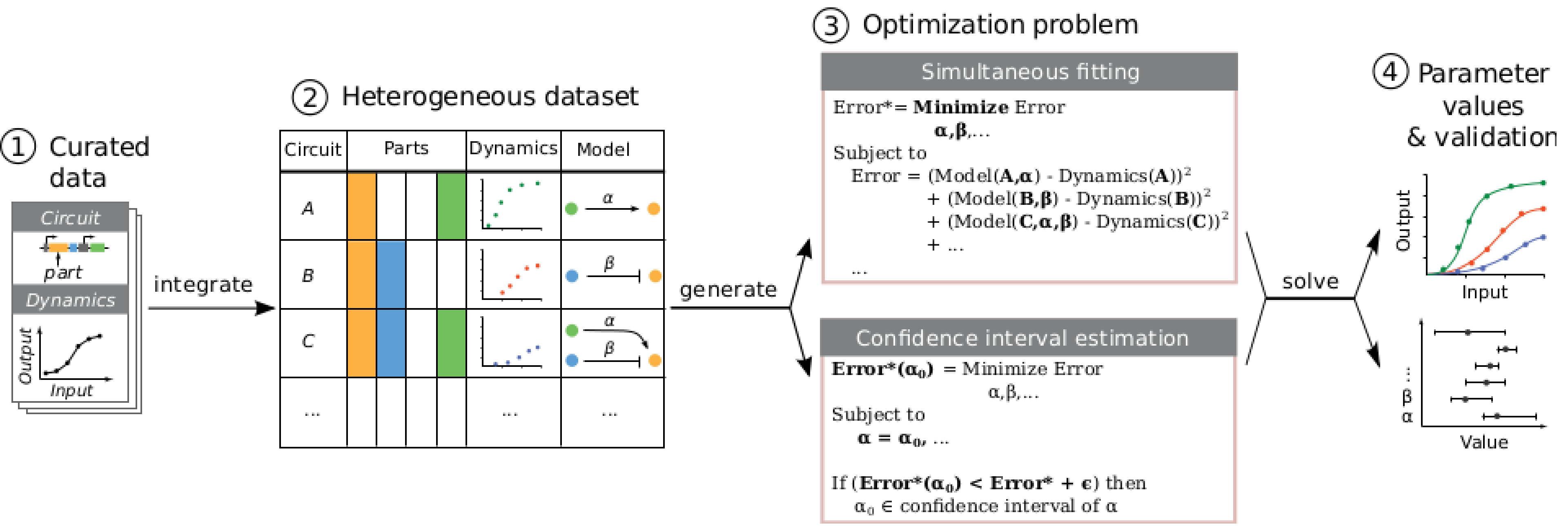
Why do we have multiple fitting solutions?

- Parameters are not well-constrained
  - The model output change due to the change of one parameter can be compensated by the change from changing other parameters.
- Can we add more constraints for parameters?
  - Do more experiments (e.g. measure tetR)
    - **Prohibitively costly**
  - Data-driven alternative: Integrate data from overlapping circuits and infer parameters

Why do we have multiple fitting solutions?

- Parameters are not well-constrained
  - The model output change due to the change of one parameter can be compensated by the change from changing other parameters.
- Can we add more constraints for parameters?
  - Do more experiments (e.g. measure tetR)
    - **Prohibitively costly**
  - Data-driven alternative: Integrate data from overlapping circuits and infer parameters
    - **Simple**

Approach



1 Curated data
Circuit
part
Dynamics
Output
Input
integrate
2 Heterogeneous dataset
Circuit
Parts
Dynamics
Model
A
B
C
α
β
...
generate
3 Optimization problem
Simultaneous fitting
Error* = Minimize Error
α,β,...
Subject to
Error = (Model(A,α) - Dynamics(A))²
+ (Model(B,β) - Dynamics(B))²
+ (Model(C,α,β) - Dynamics(C))²
+ ...
...
Confidence interval estimation
Error*(α₀) = Minimize Error
α,β,...
Subject to
α = α₀, ...
If (Error*(α₀) < Error* + ε) then
α₀ ∈ confidence interval of α
solve
4 Parameter values & validation
Output
Input
α
β
Value

Curated data

Curated data

- From 138 publications containing gene circuits of *E. coli*

UCDAVIS
UNIVERSITY OF CALIFORNIA

Curated data

- From 138 publications containing gene circuits of *E. coli*
- Only 43 publications containing gene circuits with
  - No protein-protein interaction
  - No RNA-RNA interaction
  - No metabolism
  - No cell-cell communication
  - No recombination
  - Steady state characterization data

Curated data

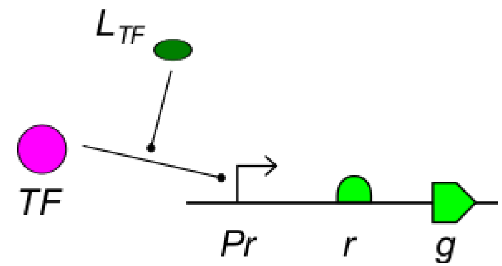
- From 138 publications containing gene circuits of *E. coli*
- Only 43 publications containing gene circuits with
  - No protein-protein interaction
  - No RNA-RNA interaction
  - No metabolism
  - No cell-cell communication
  - No recombination
  - Steady state characterization data
- Limit to only pLAC, pTET, pBAD, and constitutive promoters

Curated data

- From 138 publications containing gene circuits of *E. coli*
- Only 43 publications containing gene circuits with
  - No protein-protein interaction
  - No RNA-RNA interaction
  - No metabolism
  - No cell-cell communication
  - No recombination
  - Steady state characterization data
- Limit to only pLAC, pTET, pBAD, and constitutive promoters
  - Only 13 publications with 34 datasets

Circuit model

Translation $\mu_g = N_g \alpha_r \nu_g^{(1)}$



Transcription

$$\nu_g^{(1)} = \begin{cases} \beta_{Pr} + \dfrac{\alpha_{Pr} - \beta_{Pr}}{1 + \left(\frac{K_{Pr}}{\mu'_{TF}}\right)^{n_{Pr}}} & TF \text{ is an activator} \\ \beta_{Pr} + \dfrac{\alpha_{Pr} - \beta_{Pr}}{1 + \left(\frac{\mu'_{TF}}{K_{Pr}}\right)^{n_{Pr}}} & TF \text{ is a repressor} \end{cases}$$

Ligand binding

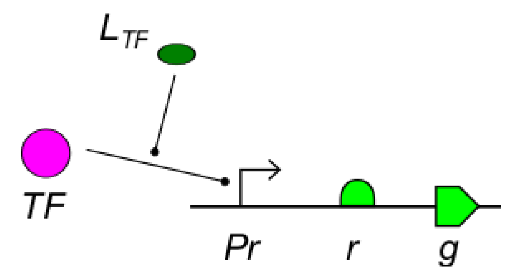
$$\mu'_{TF} = \begin{cases} \mu_{TF} & TF \text{ is non-inducible} \\ \dfrac{\mu_{TF}}{1 + (\frac{[L_{TF}]}{K_{L_{TF}}})^{n_{L_{TF}}}} & TF \text{ binds with } L_{TF} \text{ and } TF \text{ binds to } Pr \\ \dfrac{\mu_{TF}}{1 + (\frac{K_{L_{TF}}}{[L_{TF}]})^{n_{L_{TF}}}} & TF \text{ binds with } L_{TF} \text{ and } L_{TF} - TF \text{ binds to } Pr \end{cases}$$

Circuit model

Translation $$\mu_g = N_g \alpha_r \nu_g^{(1)}$$

Transcription $$\nu_g^{(1)} = \begin{cases} \beta_{Pr} + \dfrac{\alpha_{Pr} - \beta_{Pr}}{1 + \left(\frac{K_{Pr}}{\mu'_{TF}}\right)^{n_{Pr}}} & TF \text{ is an activator} \\ \beta_{Pr} + \dfrac{\alpha_{Pr} - \beta_{Pr}}{1 + \left(\frac{\mu'_{TF}}{K_{Pr}}\right)^{n_{Pr}}} & TF \text{ is a repressor} \end{cases}$$

Ligand binding $$\mu'_{TF} = \begin{cases} \mu_{TF} & TF \text{ is non-inducible} \\ \dfrac{\mu_{TF}}{1 + \left(\frac{[L_{TF}]}{K_{L_{TF}}}\right)^{n_{L_{TF}}}} & TF \text{ binds with } L_{TF} \text{ and } TF \text{ binds to } Pr \\ \dfrac{\mu_{TF}}{1 + \left(\frac{K_{L_{TF}}}{[L_{TF}]}\right)^{n_{L_{TF}}}} & TF \text{ binds with } L_{TF} \text{ and } L_{TF} - TF \text{ binds to } Pr \end{cases}$$



Consistent minimal model over all circuits

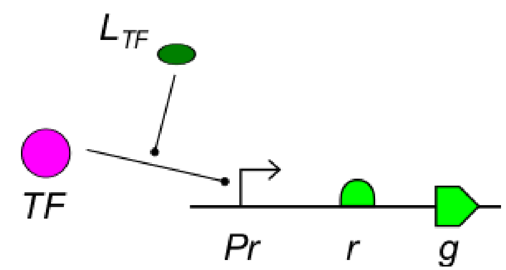
UC DAVIS
UNIVERSITY OF CALIFORNIA

Circuit model

Translation $\mu_g = N_g \alpha_r \nu_g^{(1)}$

Transcription
$$\nu_g^{(1)} = \begin{cases} \beta_{Pr} + \dfrac{\alpha_{Pr} - \beta_{Pr}}{1 + \left(\frac{K_{Pr}}{\mu'_{TF}}\right)^{n_{Pr}}} & TF \text{ is an activator} \\ \beta_{Pr} + \dfrac{\alpha_{Pr} - \beta_{Pr}}{1 + \left(\frac{\mu'_{TF}}{K_{Pr}}\right)^{n_{Pr}}} & TF \text{ is a repressor} \end{cases}$$

Ligand binding
$$\mu'_{TF} = \begin{cases} \mu_{TF} & TF \text{ is non-inducible} \\ \dfrac{\mu_{TF}}{1 + (\frac{[L_{TF}]}{K_{L_{TF}}})^{n_{L_{TF}}}} & TF \text{ binds with } L_{TF} \text{ and } TF \text{ binds to } Pr \\ \dfrac{\mu_{TF}}{1 + (\frac{K_{L_{TF}}}{[L_{TF}]})^{n_{L_{TF}}}} & TF \text{ binds with } L_{TF} \text{ and } L_{TF} - TF \text{ binds to } Pr \end{cases}$$



Consistent minimal model over all circuits

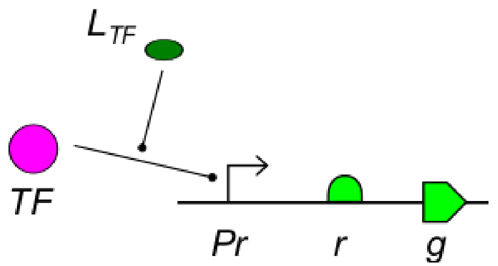
Parameter	Description	Unit
α_r	RBS strength	Relative RBS unit (RRU), normalized by BBa_B0032
$\nu_g^{(1)}$	Expression level of one copy of g	Relative promoter unit (RPU), normalized by BBa_J23101
α_{Pr}, β_{Pr}	Promoter strength & basal level	RPU
μ_g, μ_{TF}	Protein expression level of g, TF	RRU×RPU
K_{Pr}	Binding affinity	RRU×RPU
$[L_{TF}]$	Ligand concentration & dissociation constant	mM
$n_{L_{TF}}, n_{Pr}, N_g$	Hill coefficient, cooperativity & copy number	N/A

Circuit model

Translation $\mu_g = N_g \alpha_r \nu_g^{(1)}$

Transcription
$$\nu_g^{(1)} = \begin{cases} \beta_{Pr} + \dfrac{\alpha_{Pr} - \beta_{Pr}}{1 + \left(\frac{K_{Pr}}{\mu'_{TF}}\right)^{n_{Pr}}} & TF \text{ is an activator} \\ \beta_{Pr} + \dfrac{\alpha_{Pr} - \beta_{Pr}}{1 + \left(\frac{\mu'_{TF}}{K_{Pr}}\right)^{n_{Pr}}} & TF \text{ is a repressor} \end{cases}$$

Ligand binding
$$\mu'_{TF} = \begin{cases} \mu_{TF} & TF \text{ is non-inducible} \\ \dfrac{\mu_{TF}}{1 + (\frac{[L_{TF}]}{K_{L_{TF}}})^{n_{L_{TF}}}} & TF \text{ binds with } L_{TF} \text{ and } TF \text{ binds to } Pr \\ \dfrac{\mu_{TF}}{1 + (\frac{K_{L_{TF}}}{[L_{TF}]})^{n_{L_{TF}}}} & TF \text{ binds with } L_{TF} \text{ and } L_{TF} - TF \text{ binds to } Pr \end{cases}$$



Consistent minimal model over all circuits

Parameter	Description	Unit
α_r	RBS strength	Relative RBS unit (RRU), normalized by BBa_B0032
$\nu_g^{(1)}$	Expression level of one copy of g	Relative promoter unit (RPU), normalized by BBa_J23101
α_{Pr}, β_{Pr}	Promoter strength & basal level	RPU
μ_g, μ_{TF}	Protein expression level of g, TF	RRU×RPU
K_{Pr}	Binding affinity	RRU×RPU
$[L_{TF}]$	Ligand concentration & dissociation constant	mM
$n_{L_{TF}}, n_{Pr}, N_g$	Hill coefficient, cooperativity & copy number	N/A

Relative unit

UCDAVIS UNIVERSITY OF CALIFORNIA

Fitting a single model

- Log-likelihood

$$LL(D|\theta) = -\frac{1}{2}\sum_{i=1}^{n}\left(\frac{M(x_i,\theta) - y_i}{\sigma_i}\right)^2 + const$$

UCDAVIS
UNIVERSITY OF CALIFORNIA

Fitting a single model

- Log-likelihood

Exp. data

$$LL(D|\theta) = -\frac{1}{2}\sum_{i=1}^{n}\left(\frac{M(x_i,\theta) - y_i}{\sigma_i}\right)^2 + const$$

UCDAVIS
UNIVERSITY OF CALIFORNIA

Fitting a single model

- Log-likelihood

Exp. data **parameter set**

$$LL(D|\theta) = -\frac{1}{2}\sum_{i=1}^{n}\left(\frac{M(x_i,\theta) - y_i}{\sigma_i}\right)^2 + const$$

UCDAVIS UNIVERSITY OF CALIFORNIA

Fitting a single model

- Log-likelihood

Exp. data **parameter set** **# data point**

$$LL(D|\theta) = -\frac{1}{2}\sum_{i=1}^{n}\left(\frac{M(x_i,\theta) - y_i}{\sigma_i}\right)^2 + const$$

UCDAVIS
UNIVERSITY OF CALIFORNIA

Fitting a single model

- Log-likelihood

Exp. data parameter set # data point

model

$$LL(D|\theta) = -\frac{1}{2}\sum_{i=1}^{n}\left(\frac{M(x_i,\theta)-y_i}{\sigma_i}\right)^2 + const$$

Fitting a single model

- Log-likelihood

Exp. data **parameter set** **# data point**

model **Exp. input**

$$LL(D|\theta) = -\frac{1}{2}\sum_{i=1}^{n}\left(\frac{M(x_i,\theta) - y_i}{\sigma_i}\right)^2 + const$$

Fitting a single model

- Log-likelihood

Exp. data **parameter set** **# data point**

model **Exp. input** **Exp. output**

$$LL(D|\theta) = -\frac{1}{2}\sum_{i=1}^{n}\left(\frac{M(x_i,\theta) - y_i}{\sigma_i}\right)^2 + const$$

Fitting a single model

- Log-likelihood

Exp. data **parameter set** **# data point**

model **Exp. input** **Exp. output**

$$LL(D|\theta) = -\frac{1}{2}\sum_{i=1}^{n}\left(\frac{M(x_i,\theta) - y_i}{\sigma_i}\right)^2 + const$$

Exp. output standard error

UCDAVIS
UNIVERSITY OF CALIFORNIA

Fitting a single model

- Log-likelihood

Exp. data **parameter set** **# data point** **model** **Exp. input** **Exp. output**

$$LL(D|\theta) = -\frac{1}{2}\sum_{i=1}^{n}\left(\frac{M(x_i,\theta) - y_i}{\sigma_i}\right)^2 + const$$

Exp. output standard error

- Maximum log-likelihood

$$\theta^* = \underset{\theta}{argmax}\ LL(D|\theta)$$

UCDAVIS
UNIVERSITY OF CALIFORNIA

Confidence interval

UCDAVIS
UNIVERSITY OF CALIFORNIA

Confidence interval

- Measuring the uncertainty of a parameter

UC DAVIS
UNIVERSITY OF CALIFORNIA

Confidence interval

- Measuring the uncertainty of a parameter
  - Longer confidence interval (CI) $\rightarrow$ more uncertain

Confidence interval

- Measuring the uncertainty of a parameter
  - Longer confidence interval (CI) $\rightarrow$ more uncertain
- Estimate through the profile log-likelihood

$$CI_\alpha(\theta_i) = \left\{ x \mid \max_{\theta|\theta_i=x} LL(D|\theta) \leq \max_\theta LL(D|\theta) - \Delta(\alpha) \right\}$$

UCDAVIS
UNIVERSITY OF CALIFORNIA

Confidence interval

- Measuring the uncertainty of a parameter
  - Longer confidence interval (CI) $\rightarrow$ more uncertain
- Estimate through the profile log-likelihood

$$CI_\alpha(\theta_i) = \left\{ x \mid \max_{\theta|\theta_i=x} LL(D|\theta) \leq \max_{\theta} LL(D|\theta) - \Delta(\alpha) \right\}$$

One parameter

UCDAVIS
UNIVERSITY OF CALIFORNIA

Confidence interval

- Measuring the uncertainty of a parameter
  - Longer confidence interval (CI) $\rightarrow$ more uncertain
- Estimate through the profile log-likelihood

$$CI_\alpha(\theta_i) = \{x \mid \max_{\theta|\theta_i=x} LL(D|\theta) \leq \max_{\theta} LL(D|\theta) - \Delta(\alpha)\}$$

One parameter

Threshold

Fitting multiple models

Fitting multiple models

- Independent fitting
  - Fit each model independently

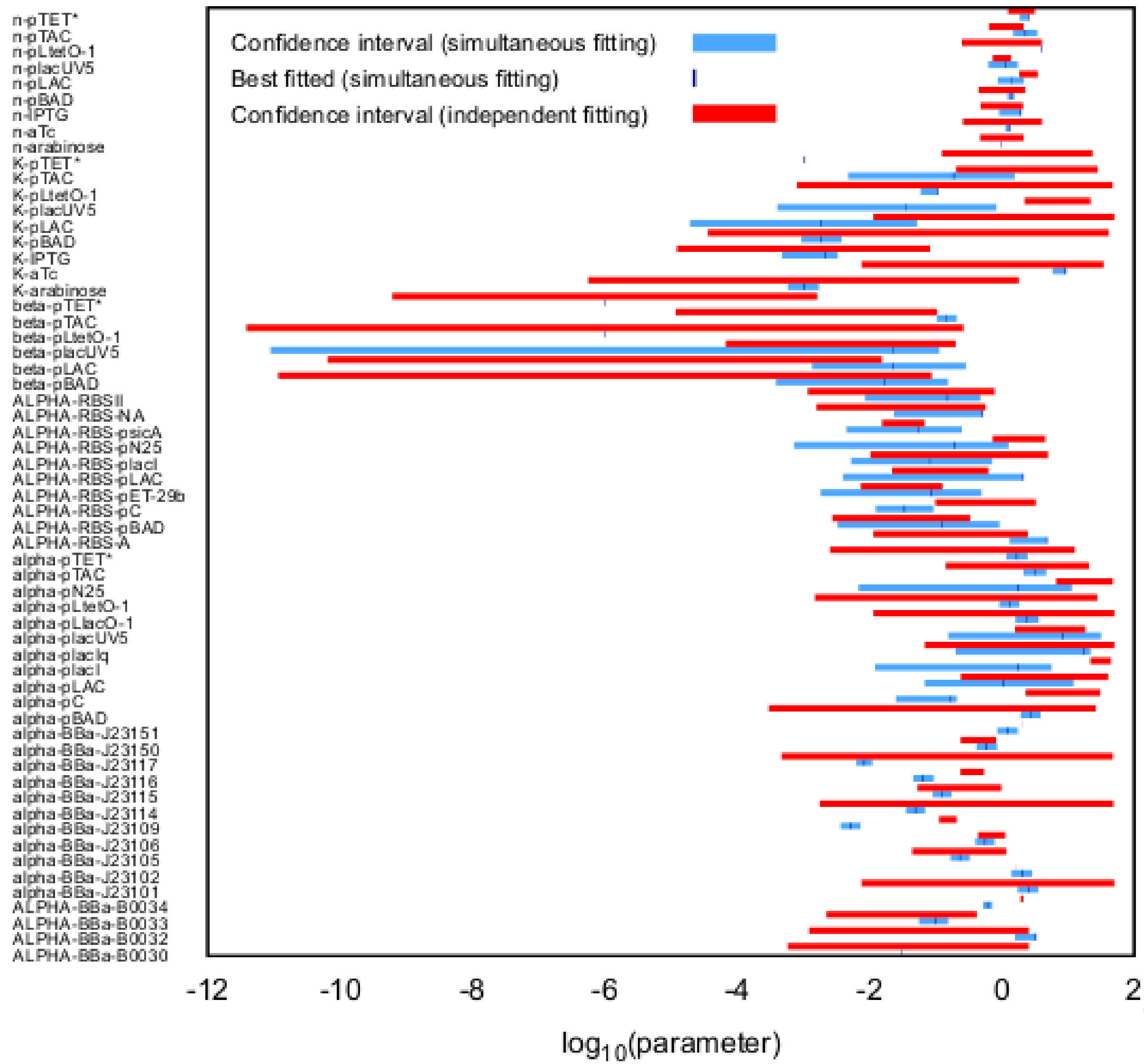
Fitting multiple models

- Independent fitting
  - Fit each model independently
- Simultaneous fitting
  - Fit all models at the same time

Fitting multiple models

- Independent fitting
  - Fit each model independently
- Simultaneous fitting
  - Fit all models at the same time
- Sequential fitting
  - Fit each model, one by one
  - The results from the former fittings are used in the latter fittings

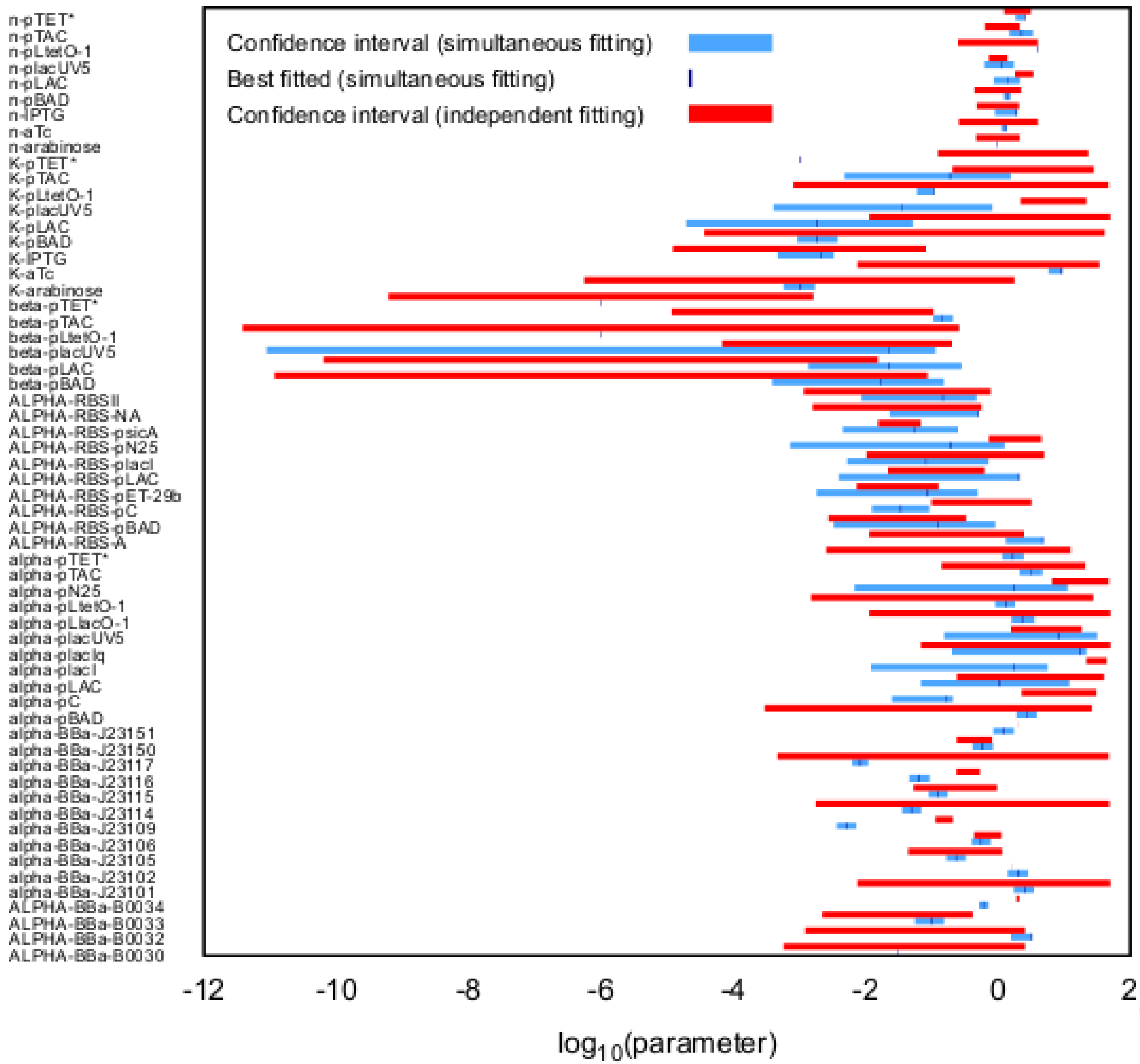
Simultaneous fitting vs independent fitting



UC DAVIS
UNIVERSITY OF CALIFORNIA

Simultaneous fitting vs independent fitting

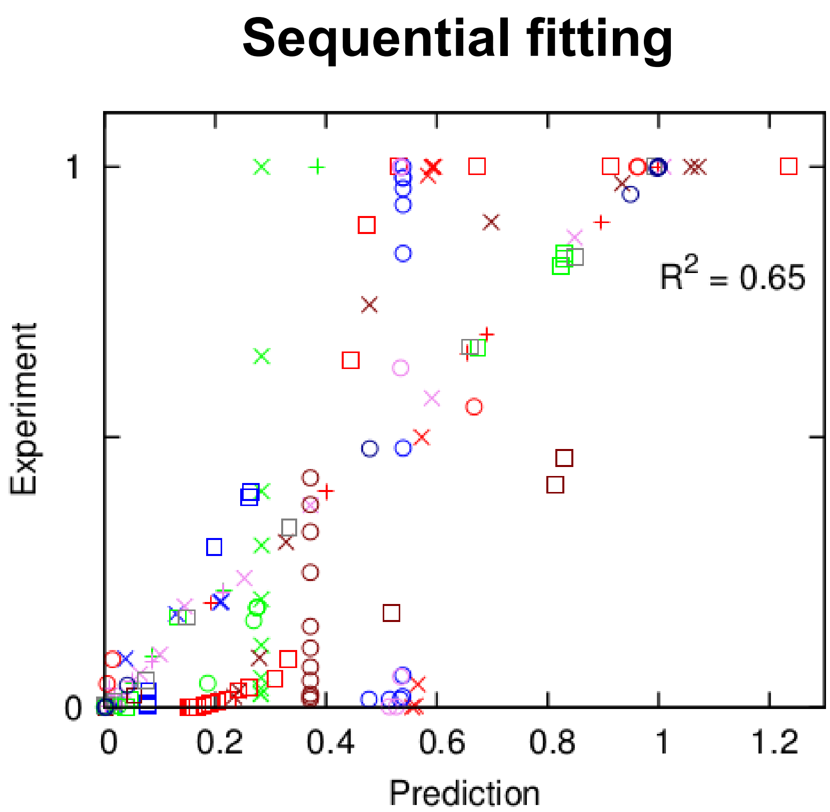
Confidence interval length (log-scale) reduces 19% in average

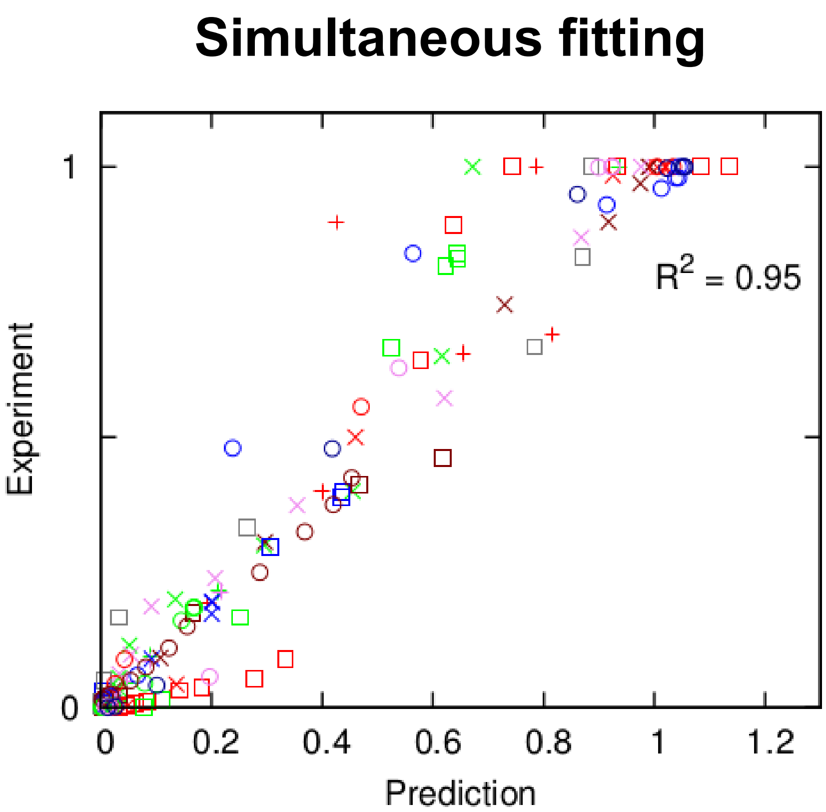


UCDAVIS
UNIVERSITY OF CALIFORNIA

Simultaneous fitting vs sequential fitting

Simultaneous fitting vs sequential fitting

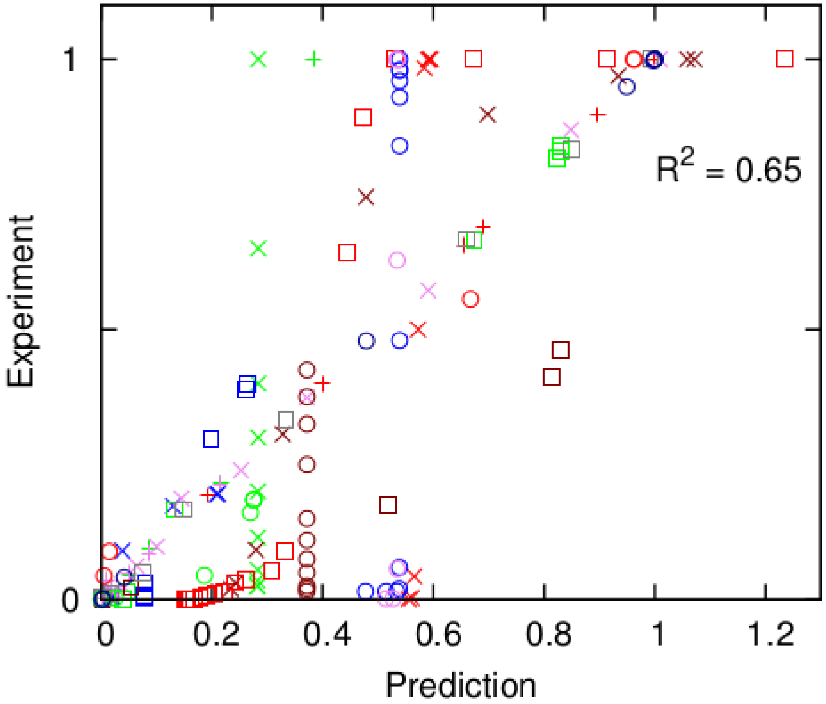




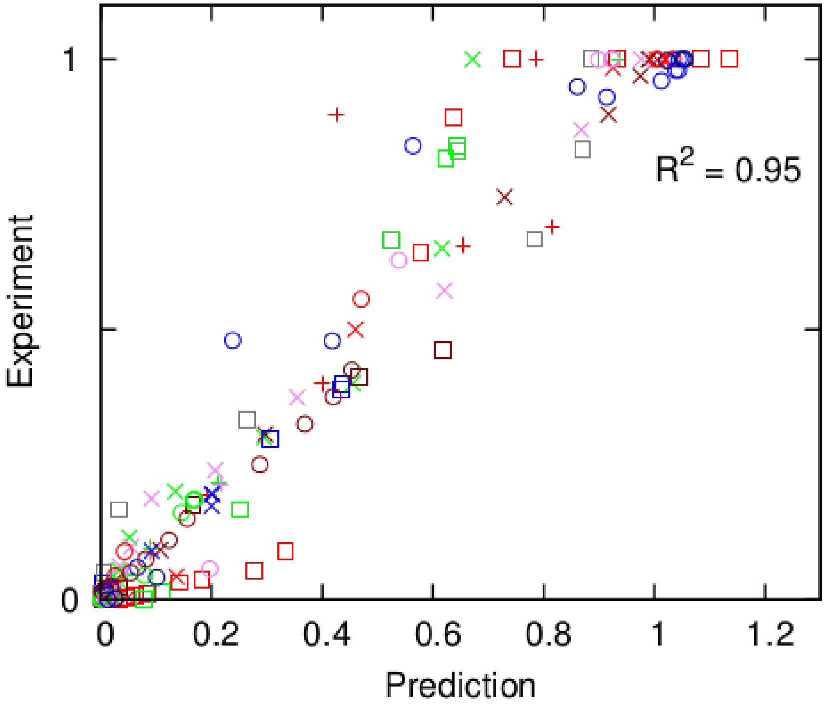
UC DAVIS
UNIVERSITY OF CALIFORNIA

Simultaneous fitting vs sequential fitting

Sequential fitting



Simultaneous fitting



Error reduction: R^2 = 0.95 vs R^2 = 0.65

UCDAVIS
UNIVERSITY OF CALIFORNIA

Results

- Data integration & simultaneous fitting

Results

- Data integration & simultaneous fitting
  - CI reduction: 19%

UCDAVIS
UNIVERSITY OF CALIFORNIA

Results

- Data integration & simultaneous fitting
  - CI reduction: 19%
  - Error reduction: $R^2 = 0.95$ vs $R^2 = 0.65$

Results

- Data integration & simultaneous fitting
  - CI reduction: 19%
  - Error reduction: $R^2 = 0.95$ vs $R^2 = 0.65$
  - Running time: 5x increase

Currently working on

Currently working on

- Model extensions

Currently working on

- Model extensions
  - RNA-RNA interactions

Currently working on

- Model extensions
  - RNA-RNA interactions
  - Protein-protein interactions

UCDAVIS
UNIVERSITY OF CALIFORNIA

Currently working on

- Model extensions
  - RNA-RNA interactions
  - Protein-protein interactions
  - Integration with genome-scale model

Currently working on

- Model extensions
  - RNA-RNA interactions
  - Protein-protein interactions
  - Integration with genome-scale model
- Online resource for dataset and inference tool

Currently working on

- Model extensions
  - RNA-RNA interactions
  - Protein-protein interactions
  - Integration with genome-scale model
- Online resource for dataset and inference tool
  - SBOL: Need an extension for specifying experimental data

Currently working on

- Model extensions
  - RNA-RNA interactions
  - Protein-protein interactions
  - Integration with genome-scale model
- Online resource for dataset and inference tool
  - SBOL: Need an extension for specifying experimental data
- Experimental validation

Acknowledgments

Tagkopoulos lab members:

- Violeta Zorraquino
- Navneet Rai
- Minseung Kim
- Ameen Eetemadi
- Xiaokang Wang
- Beatriz Pereira

Funding

BLUE WATERS PROJECT

CITRIS
Center for Information Technology Research in the Interest of Society

UCDAVIS
UNIVERSITY OF CALIFORNIA