The Novo Nordisk Foundation Center for Biosustainability

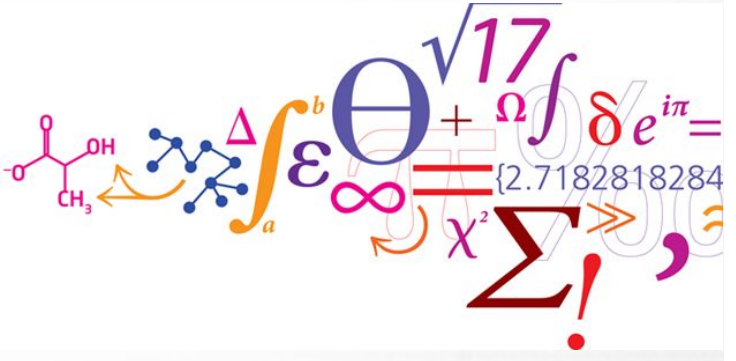
- DTU Biosustain

Making the most out of a single datapoint using Approximate Bayesian inference. Example from kinetical modeling

Denis Shepelin, PhD student

/DenisShepelin

ecol.ai



DTU Biosustain
The Novo Nordisk Foundation Center for Biosustainability

Copenhagen Bioscience PhD
novo nordisk foundation initiative

Biotechnology

Food (Beer, Dairy, ...)

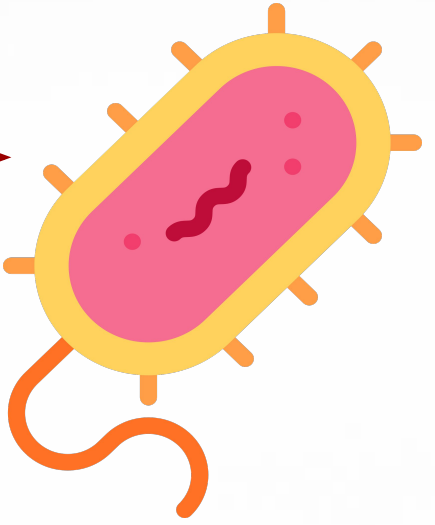
Drugs (Insulin, Herceptin, ...)

Chemicals (Plastics, fuels, ...)

Biotechnology can be used in almost any industry

Cell factories. Enzymes and fluxes

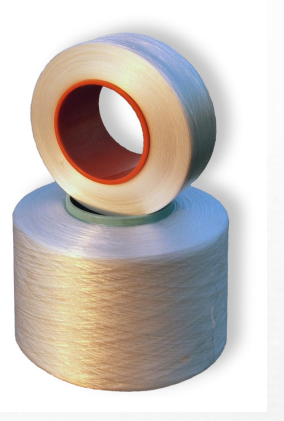
Sugar
Glycerol
Waste
...



Plastics
Fuels
Drugs
...

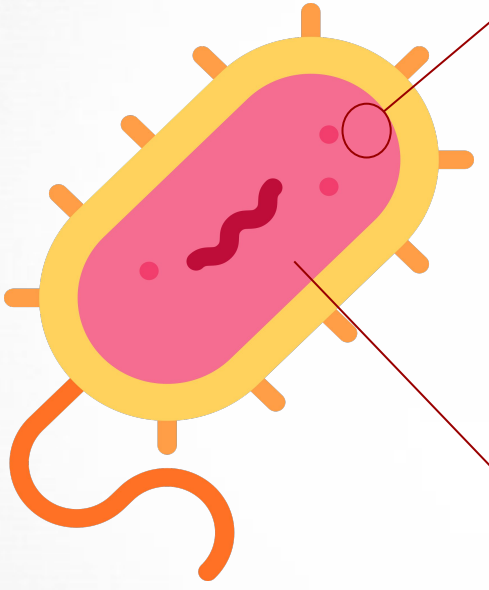
Glucose

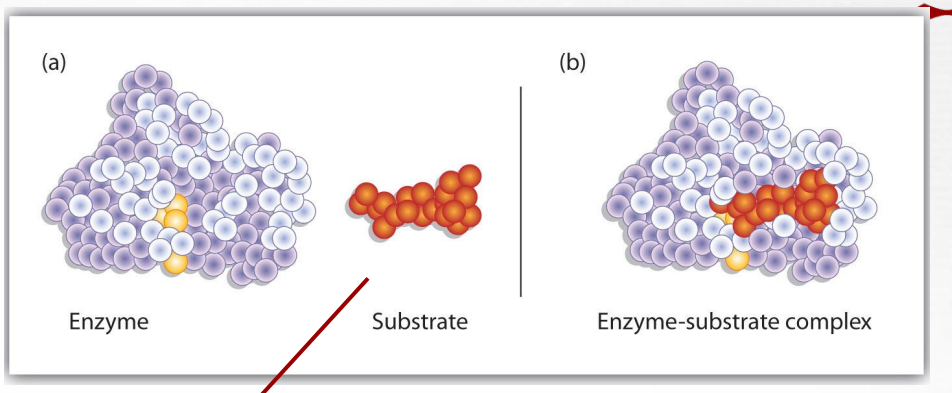
Chemical conversion



Spandex

Chemicals in cell





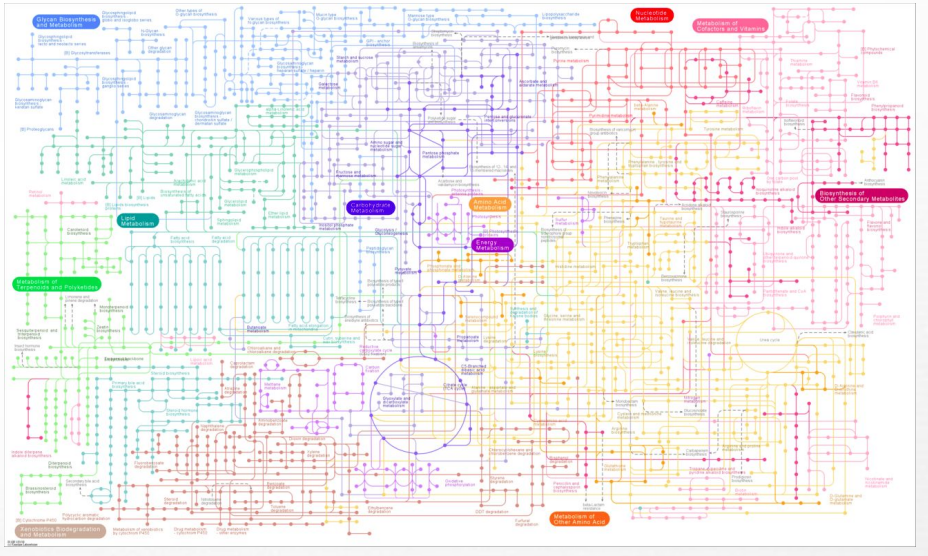
(ΔG)

Substrate

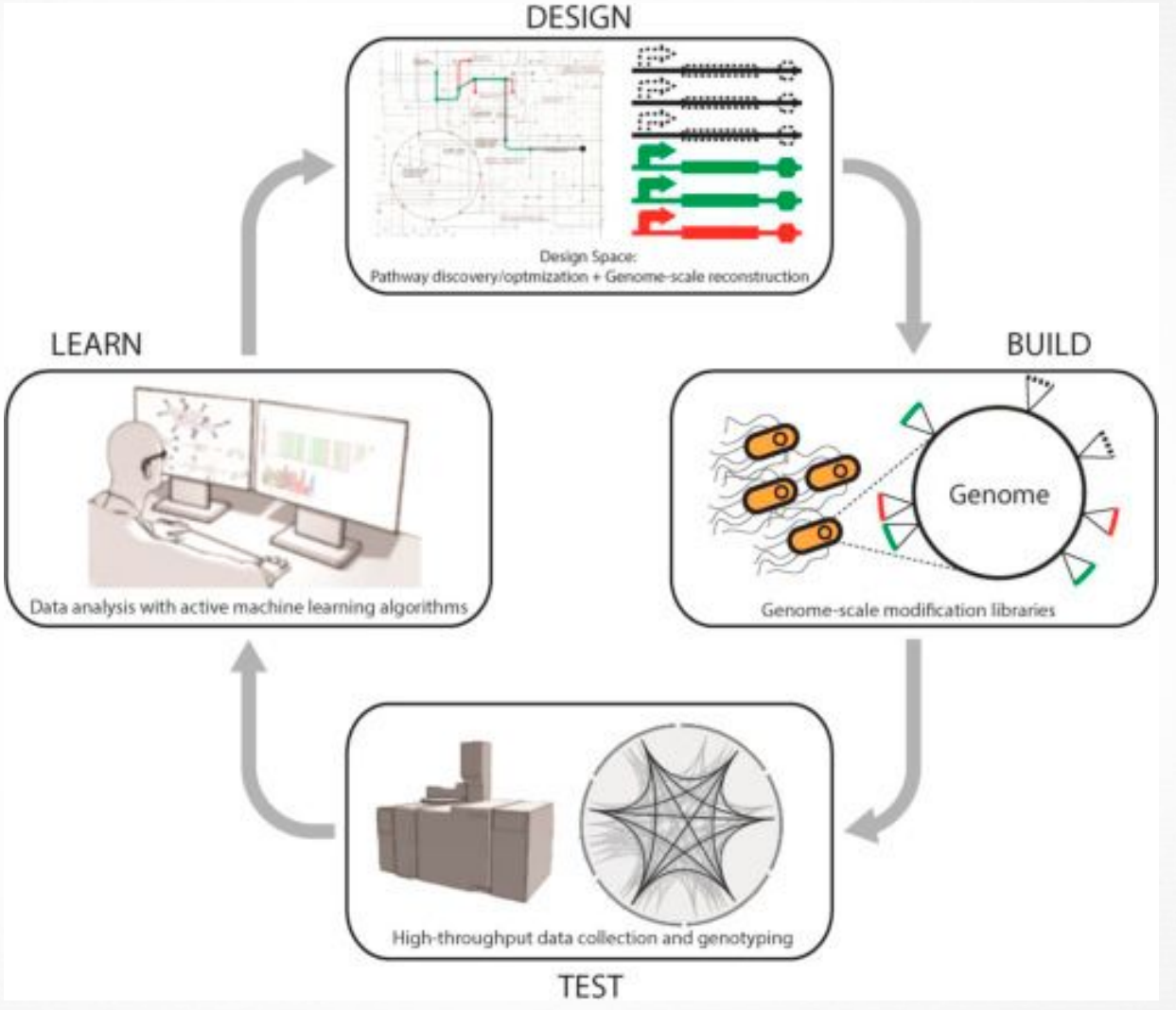
How many molecules go through reaction - **flux**

Governed by **enzymes**

Product



Biotechnology the modern way



(Un)surprisingly hard!

https://doi.org/10.1016/j.ymben.2015.09.013

Data available to biologists

We have tools to define and explore **structure** of metabolic network given organism genome - we know which reactions are there and what

Techniques to measure sets of molecules simultaneously - "-omics" technologies

1. *Metabol*omics - **abundance of chemicals** (metabolites). Usually ≈ 100s of features per sample.

2. *Prote*omics - **abundance of proteins (enzymes)**. Usually ≈ 1000s of features per sample.

3. *Flux*omics - **estimates of reaction fluxes**. Usually ≈ 100s of features per sample.

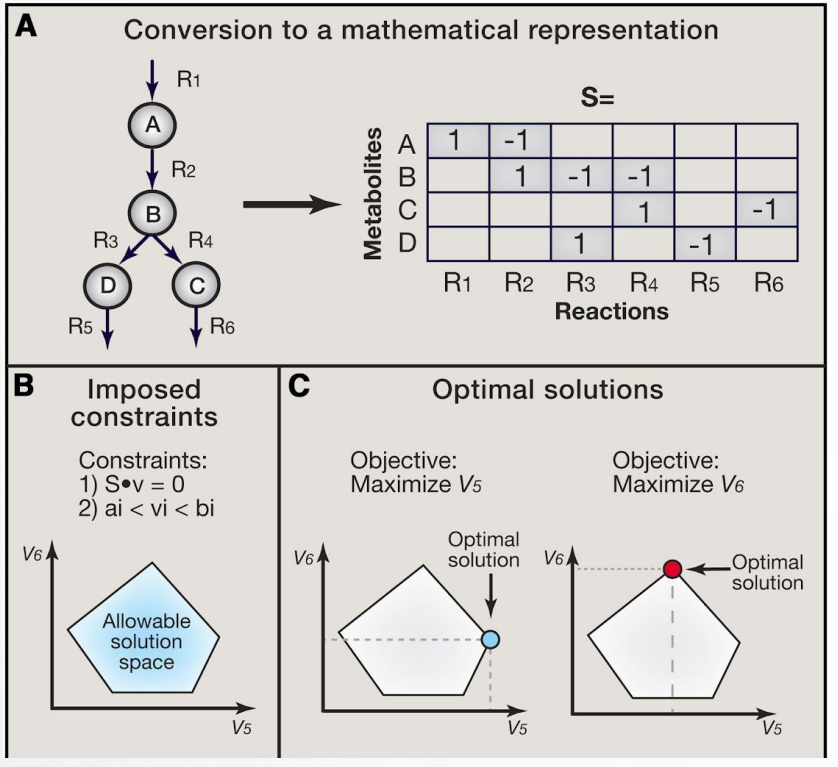
Data is noisy. Sometimes we are not sure about noise structure (Not Gaussian)

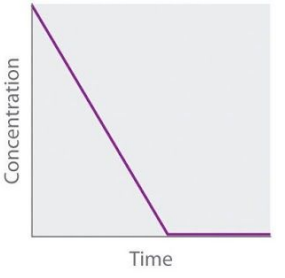
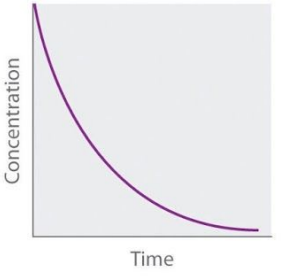
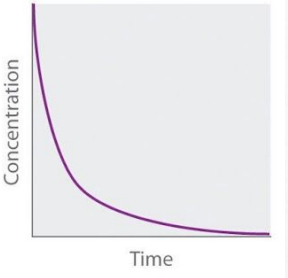
Describing metabolism. Chemical kinetics. Thermodynamics

Thermodynamics (ΔG)- possibility of reaction, **kinetics** - speed of reaction

Metabolic network structure as transport problem
Linear programming problem

Chemical transformation as kinetical equations
System of ODEs



Zeroth Order	First Order	Second Order
$\text{Rate} = -\frac{\Delta[A]}{\Delta t} = k$	$\text{Rate} = -\frac{\Delta[A]}{\Delta t} = k[A]$	$\text{Rate} = -\frac{\Delta[A]}{\Delta t} = k[A]^2$
		
$[A] = [A]_0 - kt$	$[A] = [A]_0 e^{-kt}$ or $1n[A] = 1n[A]_0 - kt$	$\frac{1}{[A]} = \frac{1}{[A]_0} + kt$

Using Genome-scale Models to Predict Biological Capabilities; https://doi.org/10.1016/j.cell.2015.05.019

https://derekcarrsavvy-chemist.blogspot.dk/2016/02/reaction-kinetics-5-kinetics-and.html

Generalized Monod-Wyman-Changeux model

MWC describes chemical kinetics accounting for **many kinds of events** - is very **complex** and hard to fit

$$\frac{d\tilde{\mathbf{x}}}{dt} = \mathrm{diag}\left(\mathbf{x}^{\mathrm{ref}}\right)^{-1} \cdot \mathbf{S} \cdot \mathbf{v}\left(\tilde{\mathbf{E}}; \tilde{\mathbf{x}}; \mathbf{e}; \mathbf{R}\right)$$

$$v = \Phi_{cat} * \Psi_{reg}$$

$$v_i = \underbrace{E_i \cdot nf_{\mathrm{R}}\left(\mathbf{x}_{\mathrm{M}}; \mathbf{k}_{\mathrm{R}}\right)}_{\text{catalytic}} \cdot \underbrace{\frac{1 + \left(f_{\mathrm{T}}\left(\mathbf{x}_{\mathrm{M}}; \mathbf{k}_{\mathrm{T}}\right)/f_{\mathrm{R}}\left(\mathbf{x}_{\mathrm{M}}; \mathbf{k}_{\mathrm{R}}\right)\right) \cdot Q\left(L; \mathbf{x}_{\mathrm{M}}; \mathbf{x}_{\mathrm{E}}; \mathbf{k}_{\mathrm{R}}; \mathbf{k}_{\mathrm{T}}; \mathbf{k}_{\mathrm{E}}\right)}{1 + Q\left(L; \mathbf{x}_{\mathrm{M}}; \mathbf{x}_{\mathrm{E}}; \mathbf{k}_{\mathrm{R}}; \mathbf{k}_{\mathrm{T}}; \mathbf{k}_{\mathrm{E}}\right)}}_{\text{regulatory}}$$

Formulation, construction and analysis of kinetic models of metabolism: A review of modelling frameworks 10.1016/j.biotechadv.2017.09.005

Generalized Monod-Wyman-Changeux model

MWC describes **many kinds of events** - is very complex and hard to fit

Most of parameters we can measure!

x - concentrations of metabolites
E - abundance of enzyme (it is protein), can be in active (T) or inactive state (R)
v - reaction flux

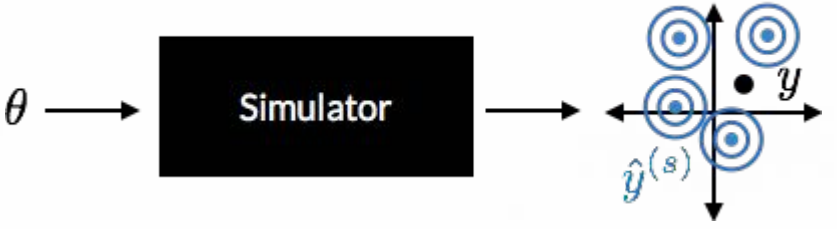
Other parameters we can sample or want to fit

k's are parameters specific to reaction (**to be fitted**)
L describes proportion of active enzyme (can be sampled) - we need **(ΔG)** here
Q is a function describing how enzymes can be activated and inactivated

Formulation, construction and analysis of kinetic models of metabolism: A review of modelling frameworks 10.1016/j.biotechadv.2017.09.005

ABC reminder

Original problem $$p(\theta \mid y) \propto p(\theta)\, p(y \mid \theta).$$

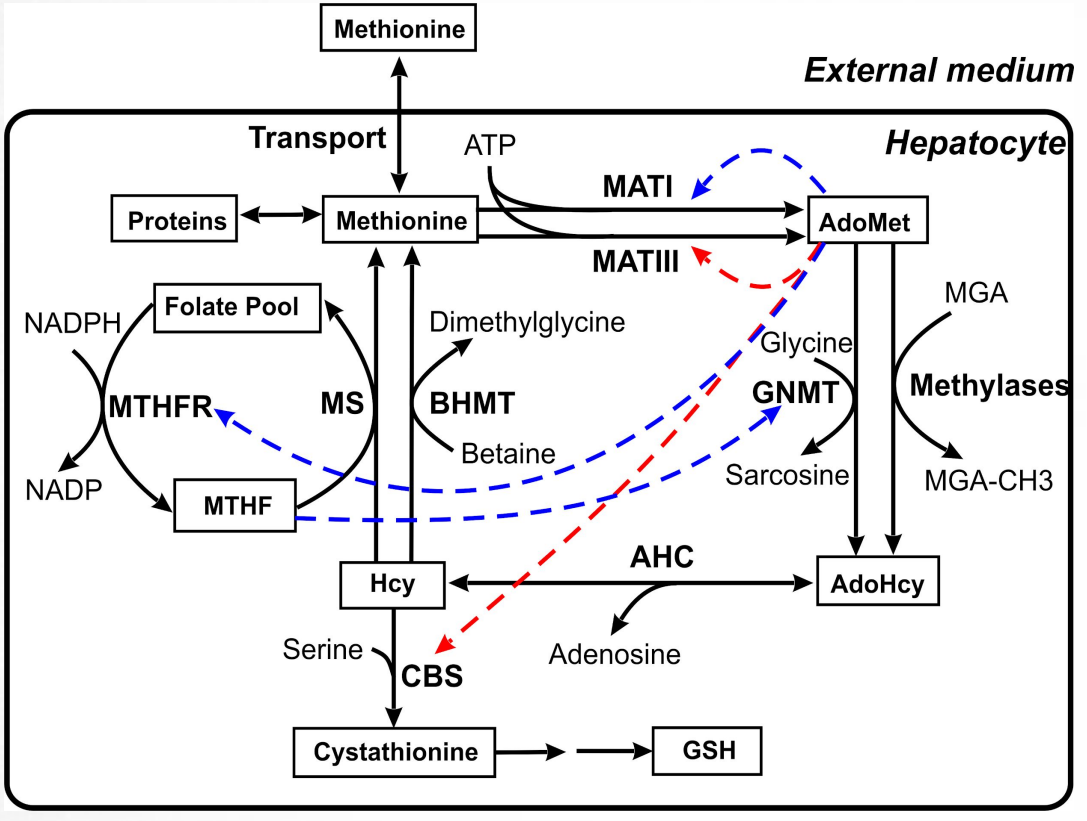


ABC “likelihood” $$p_\epsilon(y \mid \theta) = \int K_\epsilon(y, \hat{y})\, p(\hat{y} \mid \theta)\, \mathrm{d}\hat{y}$$

where K is kernel accounting for the distance between simulated sample and true data

https://casmls.github.io/general/2016/10/02/abc.html

ABC-GRASP. Methionine cycle study



Comparatively small system, has very detailed models => **good starting point**

5 ODEs + 1 algebraic equation,
72 parameters

An Allosteric Mechanism for Switching between Parallel Tracks in Mammalian Sulfur Metabolism, https://doi.org/10.1371/journal.pcbi.1000076

Case study - ABC-GRASP

θ

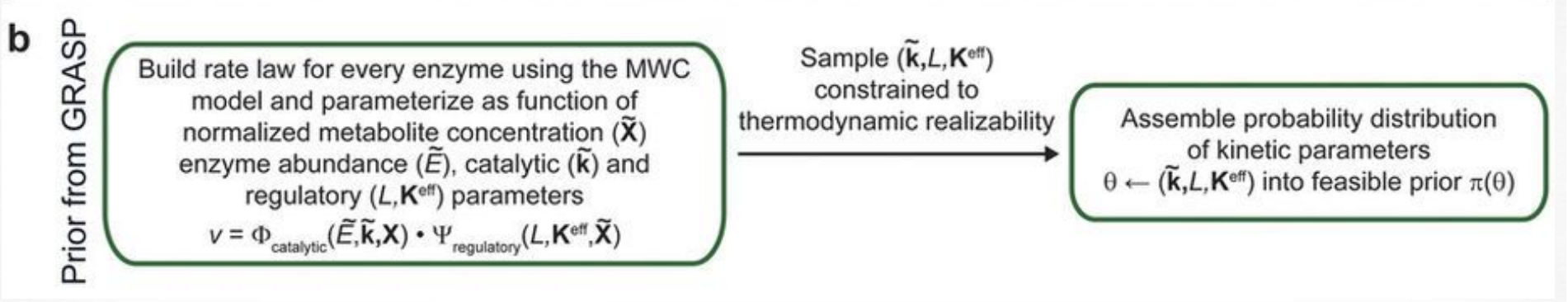
$\hat{y}$

	Structure	Data	Kinetics
INPUTS	Network structure (**S**) Thermodynamic constraints ($v_i \geq 0$) Prot. structural info.	Fluxomics ($\mathbf{v}^{ref}$,$\mathbf{v}^{exp}$) Metabolomics ($\mathbf{x}^{ref}$,$\mathbf{x}^{exp}$) Thermodynamics ($\mathbf{K}^{eq}$) Proteomics ($\mathbf{E}^{exp}$)	ER-MA and MWC $(L,\tilde{\mathbf{k}}_R,\tilde{\mathbf{k}}_T,\tilde{\mathbf{k}}_E)$
COMPUTE	Sample feasible state ($\mathbf{v}^{ref}$,$\Delta\mathbf{G}^{ref}$) Sample uniformly reversibilities ($\mathbf{R}^{ref}$), enzyme state abundances ($\mathbf{e}^{ref}$) and allosteric parameters (if necessary) Compute kinetic parameters $\tilde{\mathbf{k}}(\mathbf{R}^{ref},\mathbf{e}^{ref},\mathbf{v}^{ref},\Delta\mathbf{G}^{ref})$ and assemble reaction rates $\mathbf{v}(L,\tilde{\mathbf{k}}_R,\tilde{\mathbf{k}}_T,\tilde{\mathbf{k}}_E)$ Simulate from the model $\mathbf{v}^{sim} \leftarrow \mathbf{S}\cdot\mathbf{v}(,\mathbf{E}^{exp},\mathbf{x}^{exp}) = \mathbf{0}$, compare against $\mathbf{v}^{exp}$, and keep if they are close enough		
OUTPUT	Sample from ABC-posterior composed of a population of feasible kinetic models for prediction of metabolic states and identification of key regulatory interactions		

Construction of feasible and accurate kinetic models of metabolism: A Bayesian approach, doi:10.1038/srep29635; A General Framework for Thermodynamically Consistent Parameterization and Efficient Sampling of Enzymatic Reactions doi:10.1371/journal.pcbi.1004195

ABC scheme

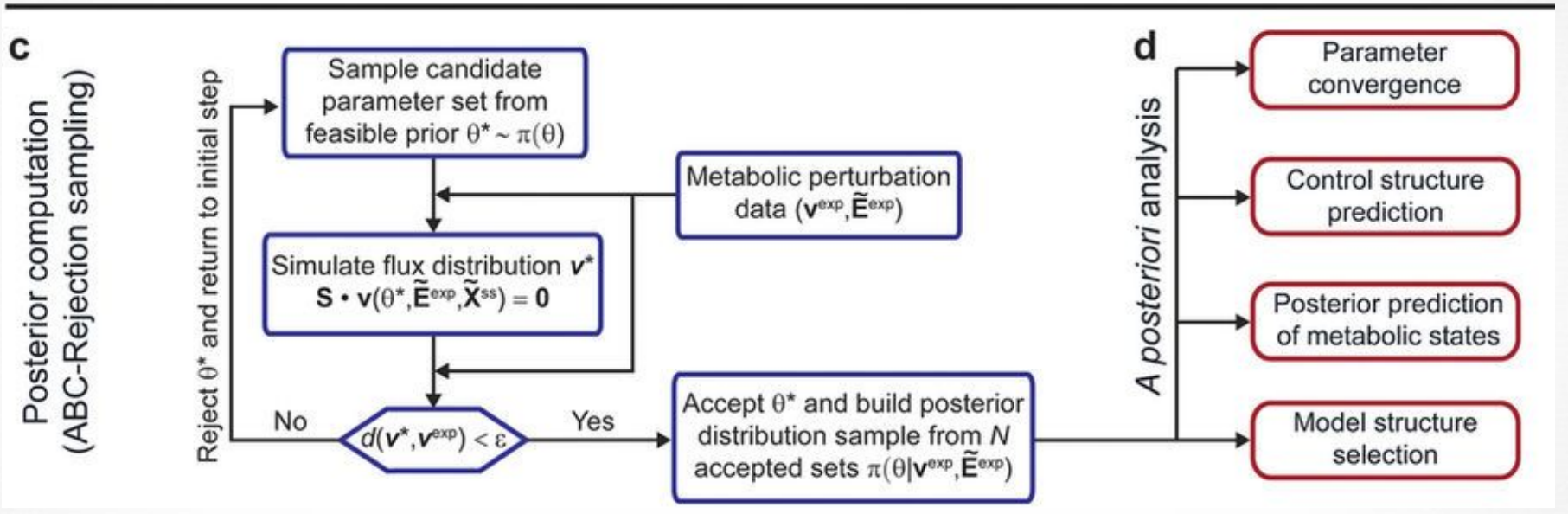
Smart choice of priors helps with sampling and defines structure.
Priors are consistent with rules of thermodynamics



$$\theta$$

ABC scheme

Parameters from the prior satisfy basic rules of chemistry =>
We save time not trying to do unrealistic simulations

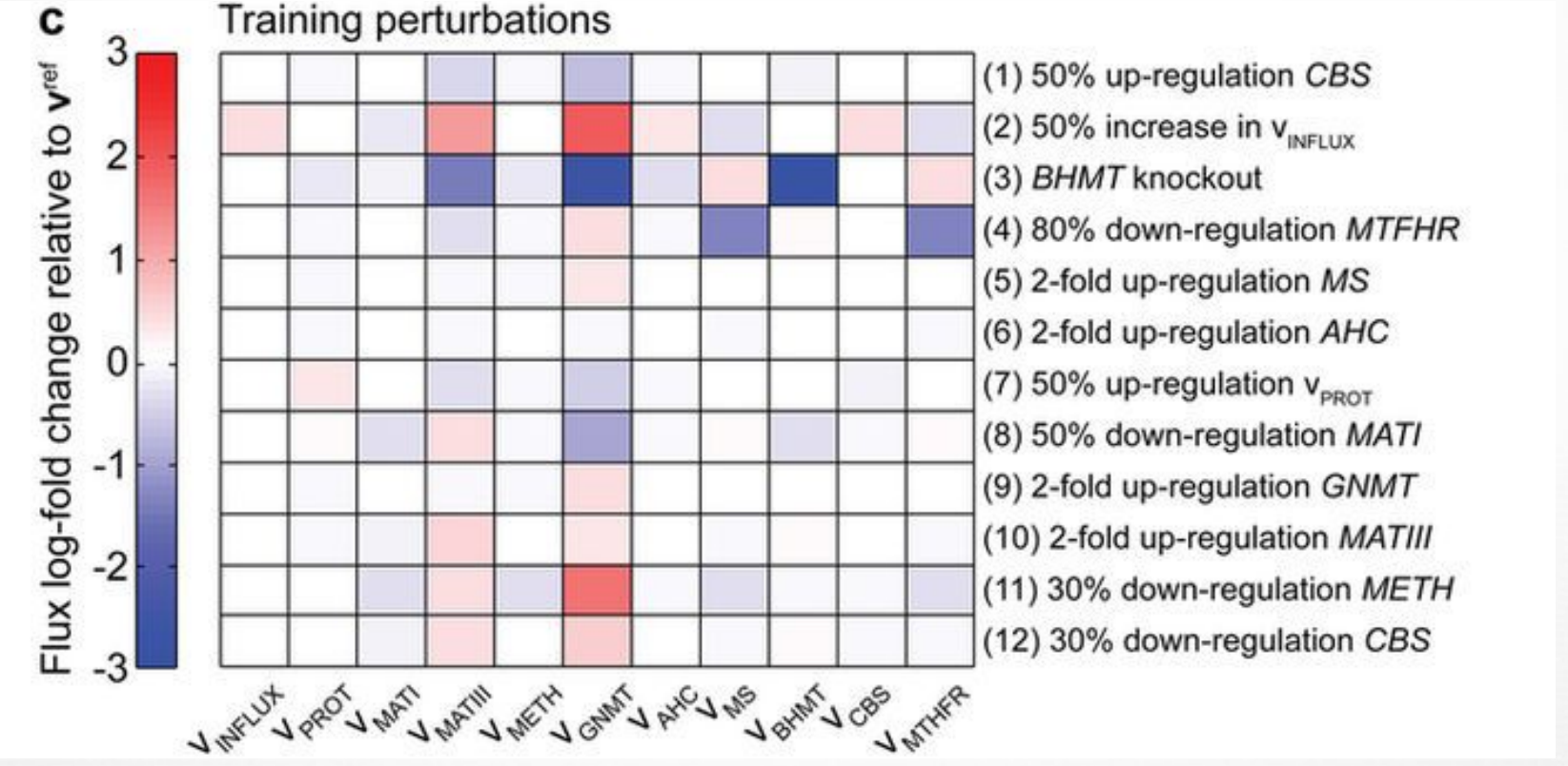


$\hat{y}$

Rejection Sampler -> Sequential Monte Carlo (experimental)

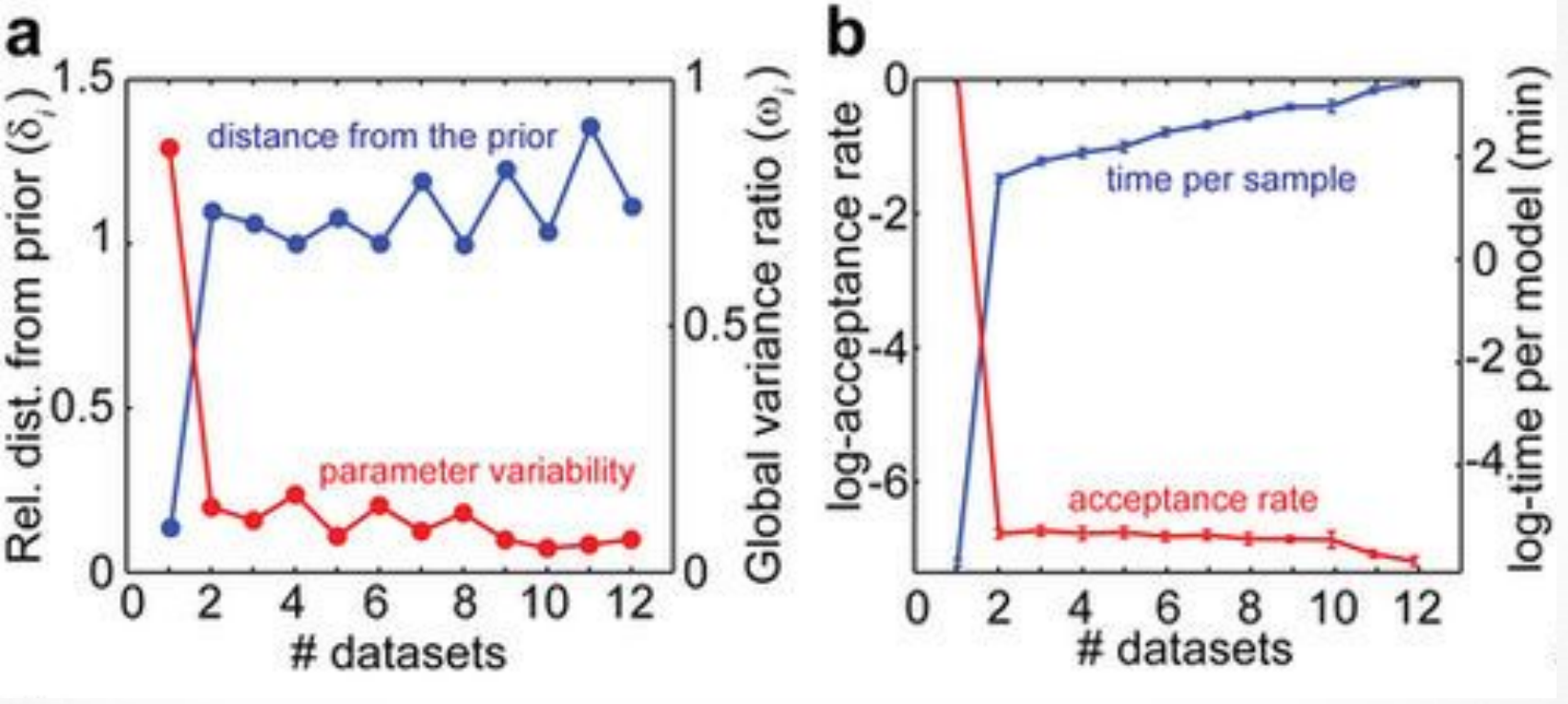
Training the model

Simulate data via published and verified model yielding 12 “samples”.
Change values of concentrations, enzyme abundancy or flux



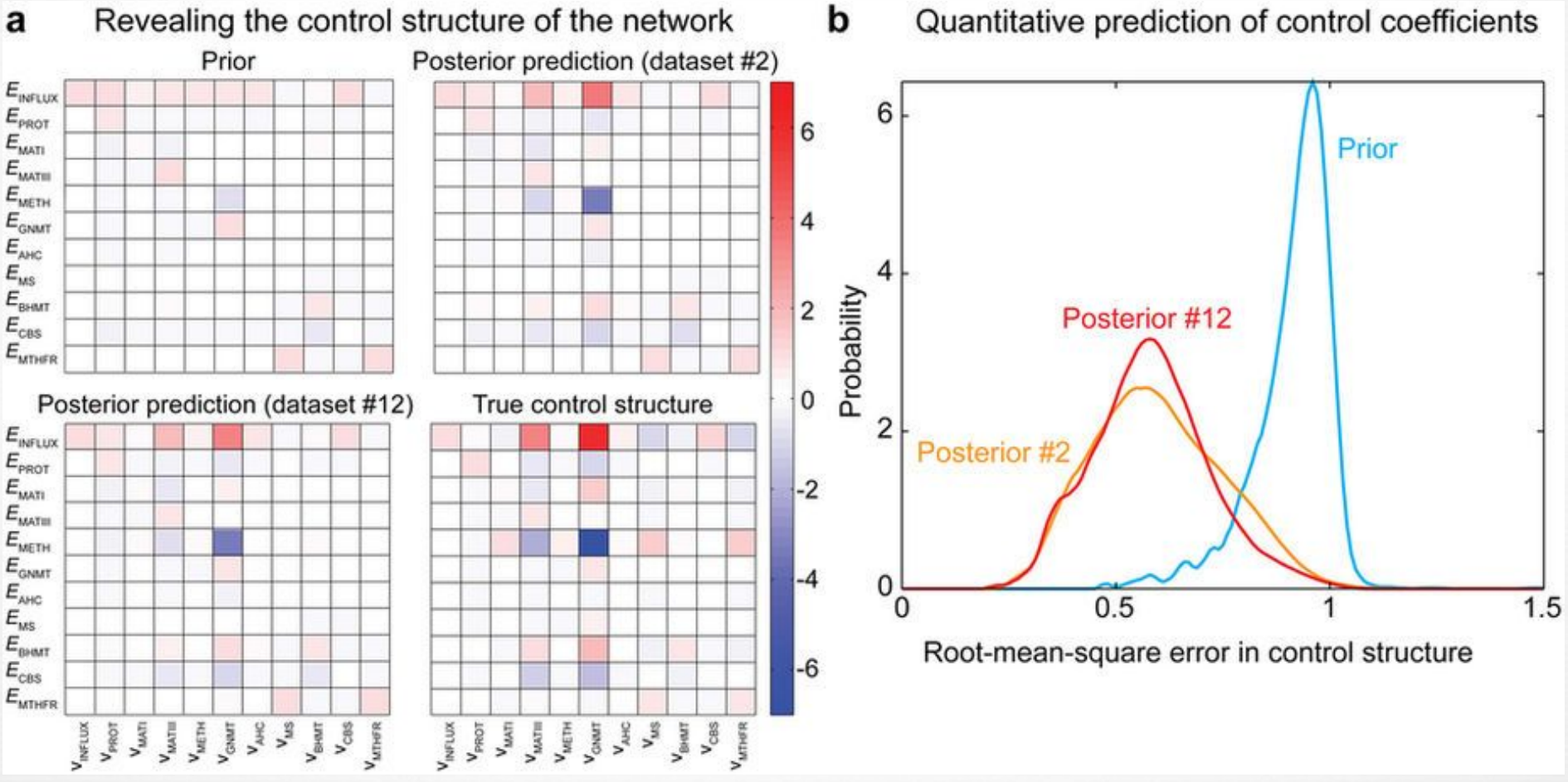
Results. Properties and Predictions

Training is fast, after two points very little changes



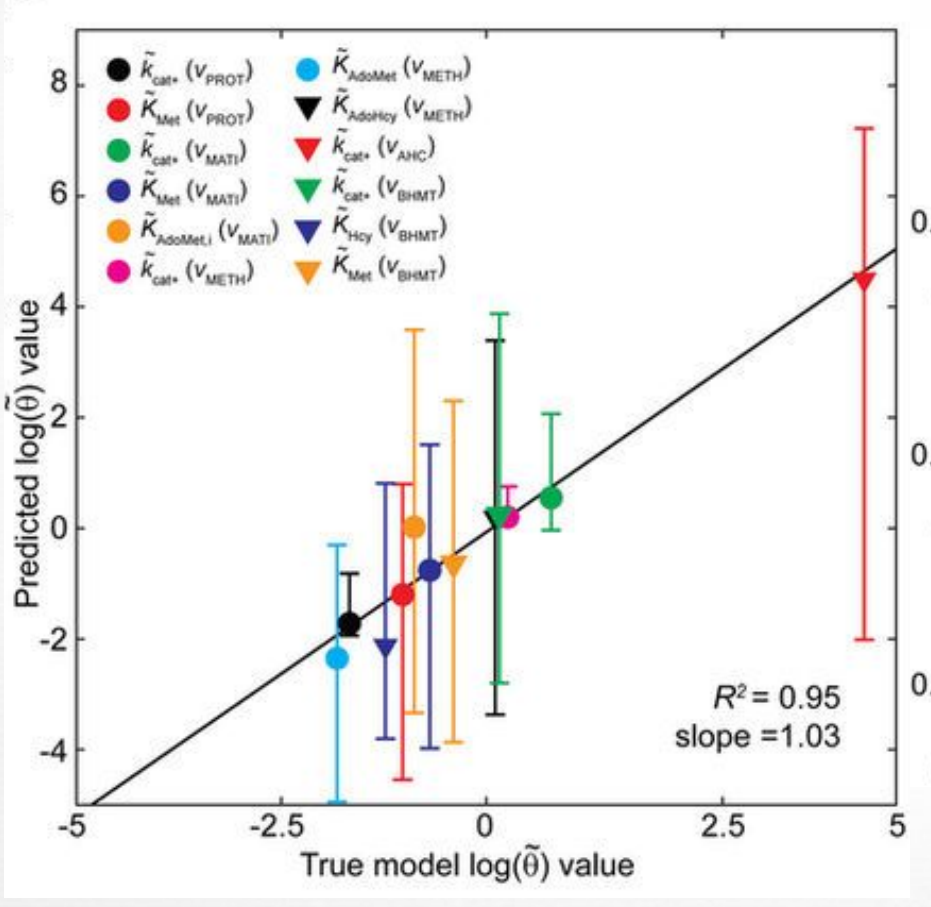
Results. Properties and Predictions

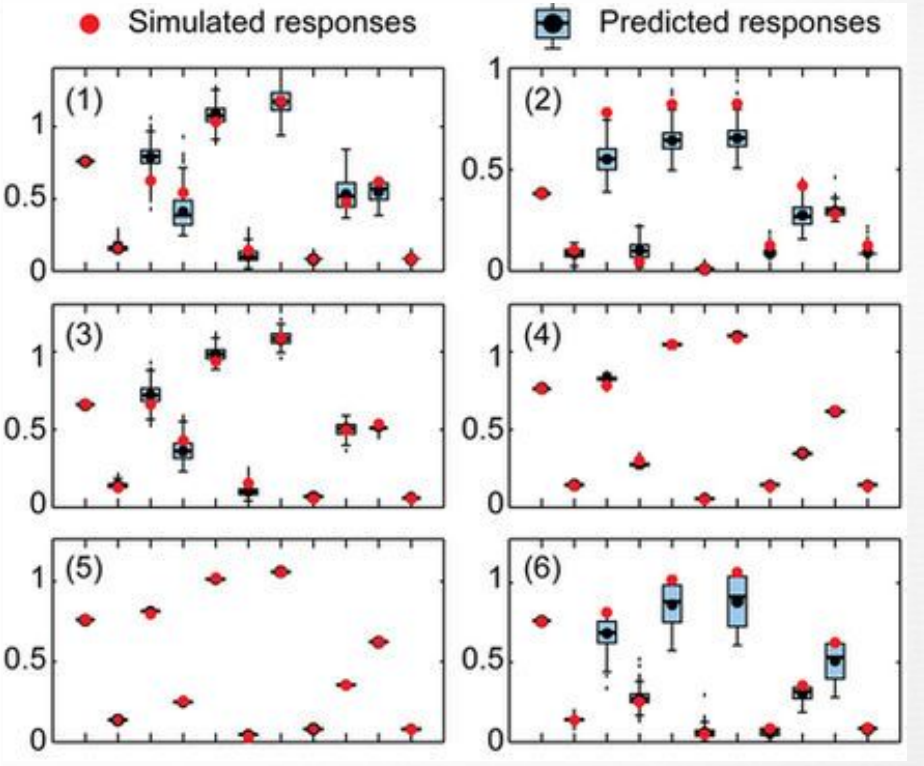
Even prior contains very valuable information. Some analyses can be performed without any data. Note that after 2 points posterior changes very slightly.



Results. Properties and Predictions

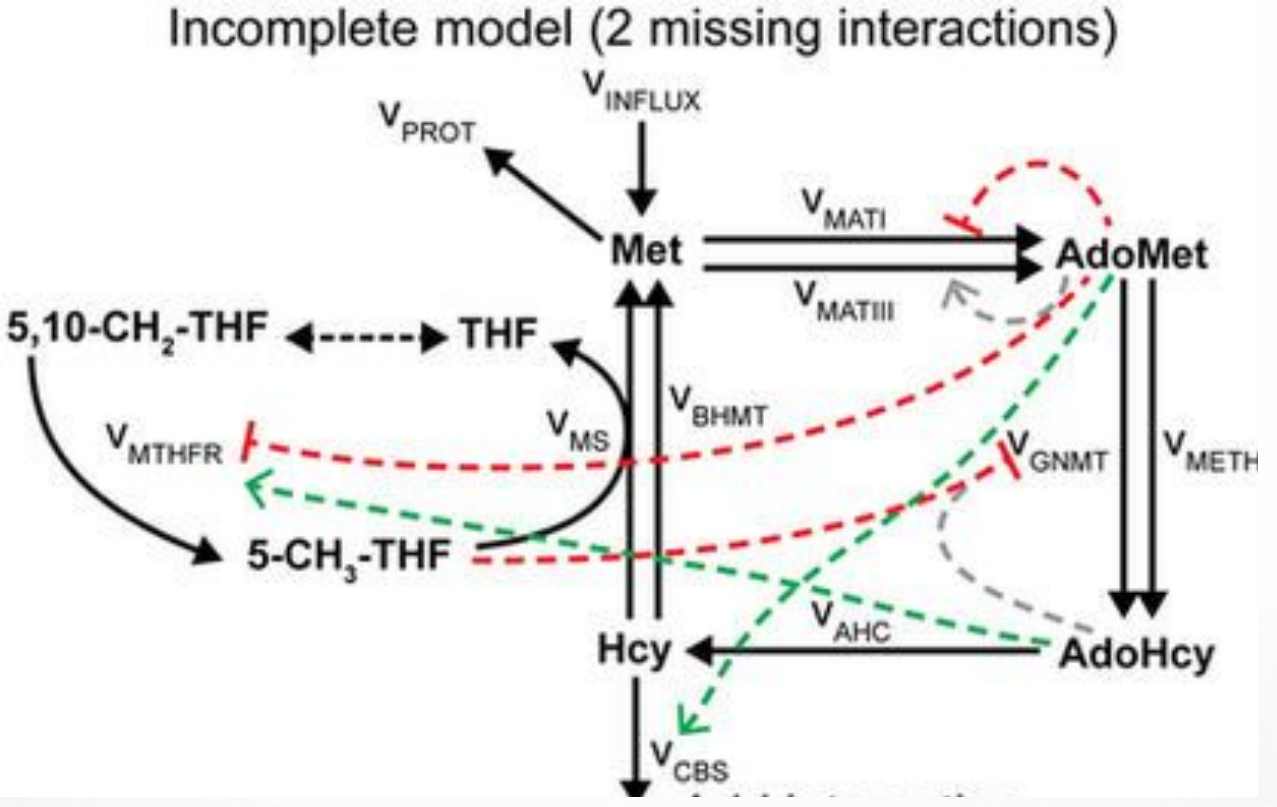
Inexact parameter fit provides accurate predictions.
We are interested in predictions!





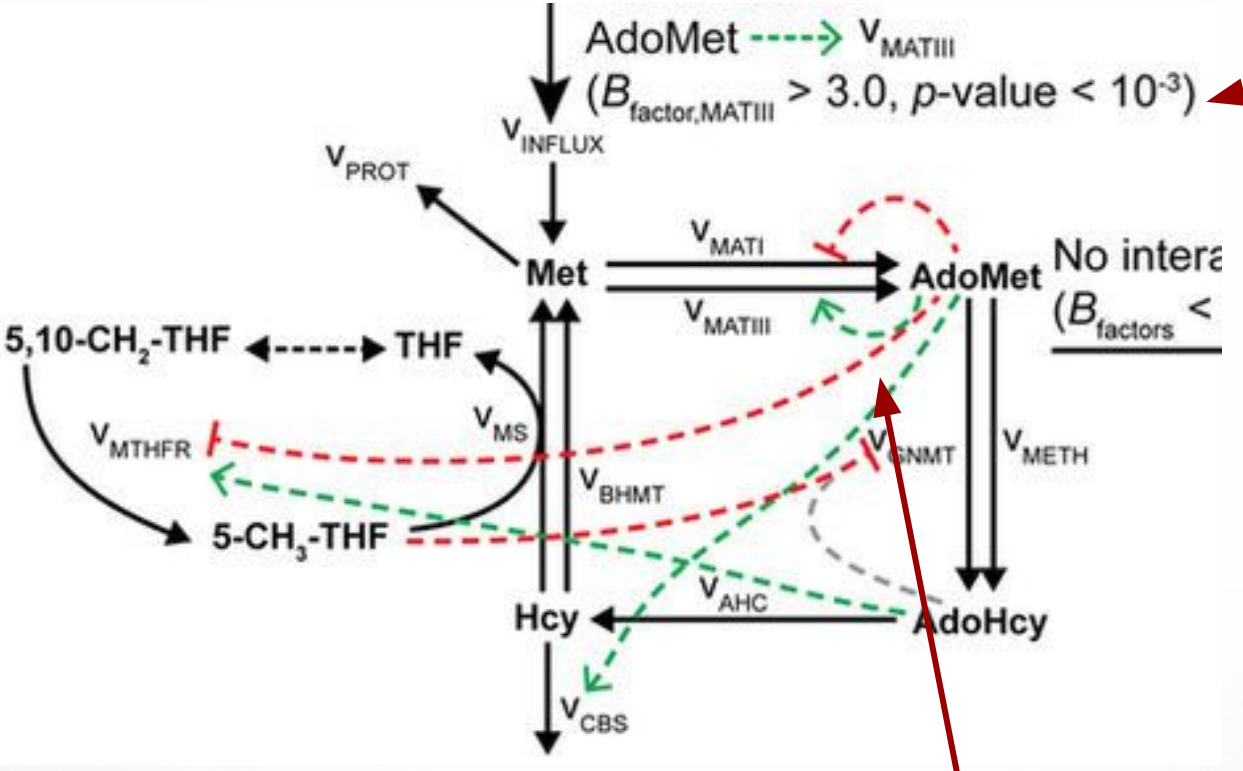
Identification of omitted rules

Some interaction between compounds and reactions are removed (grey dotted arrows).



Identification of omitted rules

Add interactions one-by-one to corrupted model. Use Bayes Factor to decide what is possible deleted interaction



BF > 3.0

Interaction recovered

Challenges

1. Computational load
2. MATLAB as environment
3. Diversity of samples - hard to control
4. How to share and communicate resulting model
5. How to scale solution to higher dimensions
6. Complicated prior (involves several linear programming routines)

Moving forward

Hamiltonian MC with information about gradients? (Graham & Storkey, 2017)

Switch from Monte-Carlo to Variational Bayes methods? (Moreno, 2016)

Probabilistic programming libraries as foundation for next-gen tools? (TensorFlow probability, Pyro, ...)

We are very happy to hear your suggestions!

Conclusions

1. We can use prior knowledge of problem structure.
2. We can use complex models within ABC framework.
3. Prediction accuracy vs parameter estimation accuracy.
4. Not all data points are equal.
5. It’s still tricky to set up and perform ABC the right way. But! there is lots of progress in the field.

ABC packages

ELFI (implements BOLFI) (Python)

pyABC from Helmholtz Centrum (Python)

ABCpy (Python)

al3c (C++)

PEITH(Θ) + abc-sysbio (Python)

abctools (R lang)

DiffEqBayes.jl (Julia)