Workflows & interactions

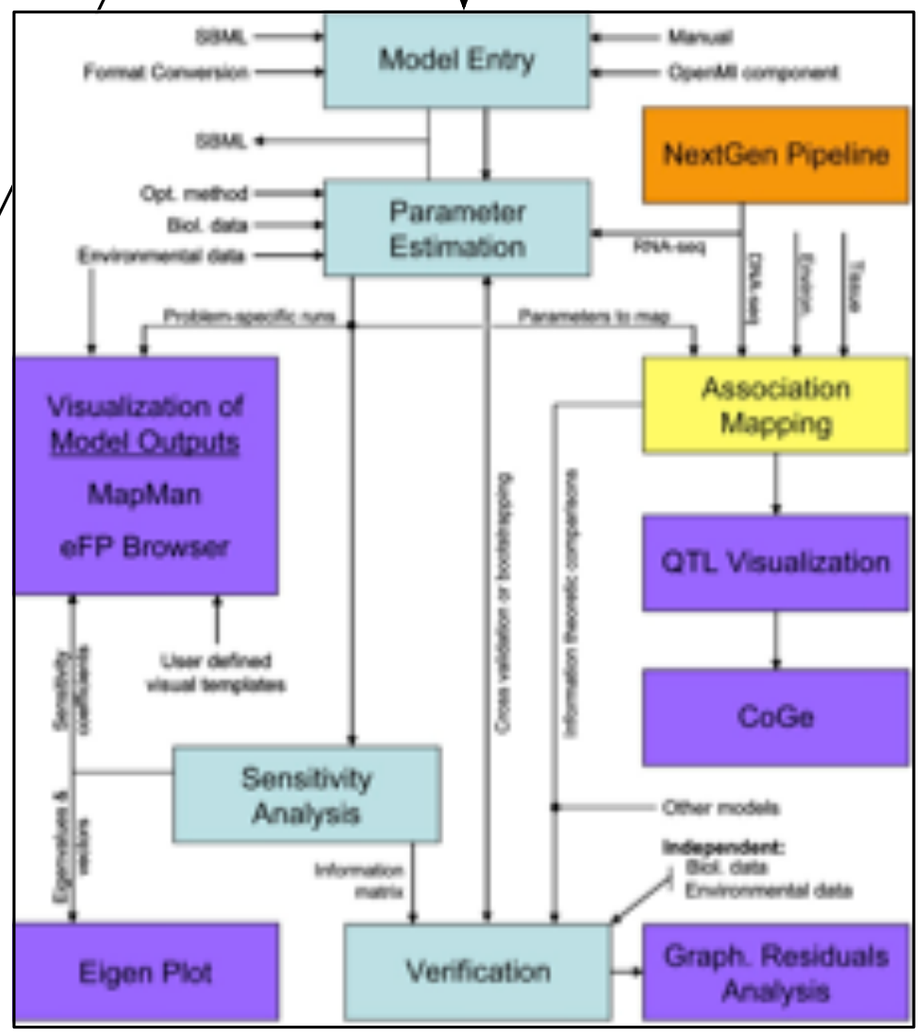
Network inference,
inferential bioinformatics
(cis-element identification,
miRNA targeting, etc.)

Data exploration & visualization,
hypothesis generation,
model curation

Data integration, orthologs,
pathway databases, interactome
databases, integration of high-
throughput and manual analyses

NextGen seq, transcriptomics,
miRNAs, ChIP-seq

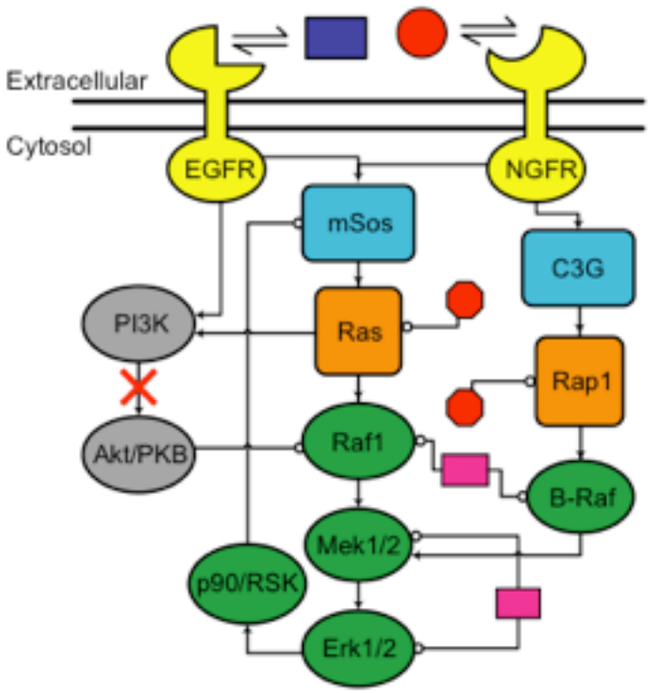
Model construction

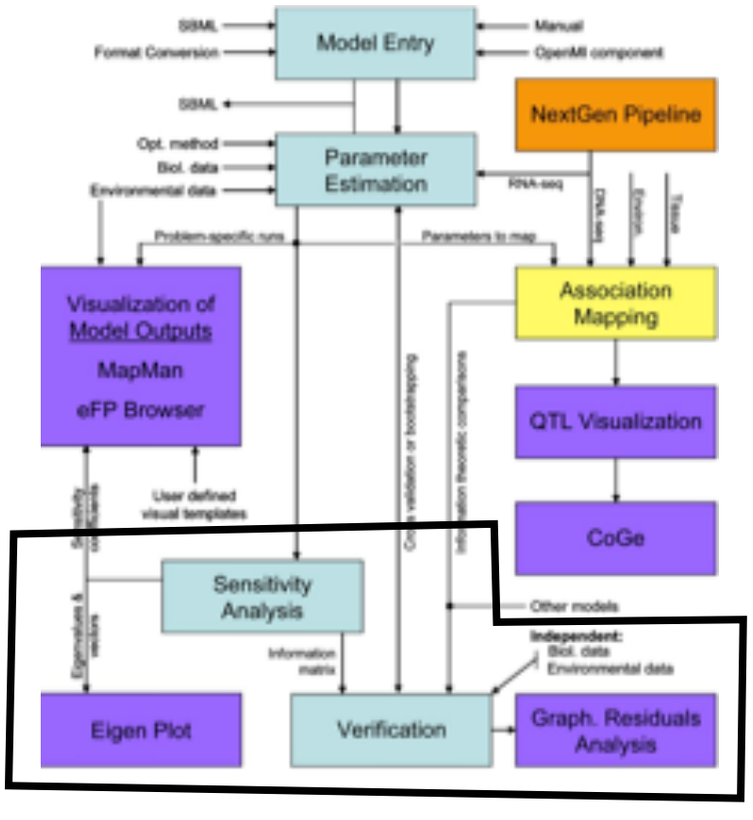


Computational pipelines

- NextGen Pipeline(s)
  - RNA-seq, DNA-seq
- cis-element prediction
  - X. Zhu: implementing various algorithms combining sequence and expression information (de novo motif identification, over-representation of candidate elements, module networks)
  - T. Mockler: ELEMENT website implementing over-representation analysis of candidate elements (PLACE db)
- miRNA identification & target prediction
  - X. Zhu: implementing tools for (i) identifying candidate miRNAs from RNA-seq data, and (ii) identifying candidate target mRNAs
- Proposal: devote iPlant resources to support pipeline refinement, coupling, and hosting as a community resource
  - where appropriate (e.g., cis-element prediction), develop common interface and tools to assess and compare results from different algorithms

Parameter sensitivities





$$\frac{d[\text{EGF}]}{dt} = -k_{\text{rbEGF}}[\text{EGF}][\text{freeEGFReceptor}] + k_{\text{ruEGF}}[\text{boundEGFR}]$$

$$\frac{d[\text{NGF}]}{dt} = -k_{\text{rbNGF}}[\text{NGF}][\text{freeNGFReceptor}] + k_{\text{ruNGF}}[\text{boundNGFR}]$$

$$\frac{d[\text{freeEGFReceptor}]}{dt} = -k_{\text{rbEGF}}[\text{EGF}][\text{freeEGFReceptor}] + k_{\text{ruEGF}}[\text{boundEGFR}]$$

$$\frac{d[\text{boundEGFR}]}{dt} = +k_{\text{rbEGF}}[\text{EGF}][\text{freeEGFReceptor}] - k_{\text{ruEGF}}[\text{boundEGFR}]$$

$$\frac{d[\text{freeNGFReceptor}]}{dt} = -k_{\text{rbNGF}}[\text{NGF}][\text{freeNGFReceptor}] + k_{\text{ruNGF}}[\text{boundNGFR}]$$

$$\frac{d[\text{boundNGFR}]}{dt} = +k_{\text{rbNGF}}[\text{NGF}][\text{freeNGFReceptor}] - k_{\text{ruNGF}}[\text{boundNGFR}]$$

$$\frac{d[\text{SosInactive}]}{dt} = -k_{\text{EGF}}[\text{boundEGFR}]\frac{[\text{SosInactive}]}{[\text{SosInactive}]+K_{\text{mEGF}}} - k_{\text{NGF}}[\text{boundNGFR}]\frac{[\text{SosInactive}]}{[\text{SosInactive}]+K_{\text{mNGF}}} + k_{\text{dSos}}[\text{P90RskActive}]\frac{[\text{SosActive}]}{[\text{SosActive}]+K_{\text{mdSos}}}$$

$$\frac{d[\text{SosActive}]}{dt} = +k_{\text{EGF}}[\text{boundEGFR}]\frac{[\text{SosInactive}]}{[\text{SosInactive}]+K_{\text{mEGF}}} + k_{\text{NGF}}[\text{boundNGFR}]\frac{[\text{SosInactive}]}{[\text{SosInactive}]+K_{\text{mNGF}}} - k_{\text{dSos}}[\text{P90RskActive}]\frac{[\text{SosActive}]}{[\text{SosActive}]+K_{\text{mdSos}}}$$

$$\frac{d[\text{P90RskInactive}]}{dt} = -k_{\text{pP90Rsk}}[\text{ErkActive}]\frac{[\text{P90RskInactive}]}{[\text{P90RskInactive}]+K_{\text{mpP90Rsk}}}$$

$$\frac{d[\text{P90RskActive}]}{dt} = +k_{\text{pP90Rsk}}[\text{ErkActive}]\frac{[\text{P90RskInactive}]}{[\text{P90RskInactive}]+K_{\text{mpP90Rsk}}}$$

$$\frac{d[\text{RasInactive}]}{dt} = -k_{\text{Sos}}[\text{SosActive}]\frac{[\text{RasInactive}]}{[\text{RasInactive}]+K_{\text{mSos}}} + k_{\text{RasGap}}[\text{RasGapActive}]\frac{[\text{RasActive}]}{[\text{RasActive}]+K_{\text{mRasGap}}}$$

$$\frac{d[\text{RasActive}]}{dt} = +k_{\text{Sos}}[\text{SosActive}]\frac{[\text{RasInactive}]}{[\text{RasInactive}]+K_{\text{mSos}}} - k_{\text{RasGap}}[\text{RasGapActive}]\frac{[\text{RasActive}]}{[\text{RasActive}]+K_{\text{mRasGap}}}$$

$$\frac{d[\text{RasGapActive}]}{dt} = 0$$

$$\frac{d[\text{Raf1Inactive}]}{dt} = -k_{\text{RasToRaf1}}[\text{RasActive}]\frac{[\text{Raf1Inactive}]}{[\text{Raf1Inactive}]+K_{\text{mRasToRaf1}}} + k_{\text{dRaf1}}[\text{Raf1PPtase}]\frac{[\text{Raf1Active}]}{[\text{Raf1Active}]+K_{\text{mdRaf1}}} + k_{\text{dRaf1ByAkt}}[\text{AktActive}]\frac{[\text{Raf1Active}]}{[\text{Raf1Active}]+K_{\text{mRaf1ByAkt}}}$$

$$\frac{d[\text{Raf1Active}]}{dt} = +k_{\text{RasToRaf1}}[\text{RasActive}]\frac{[\text{Raf1Inactive}]}{[\text{Raf1Inactive}]+K_{\text{mRasToRaf1}}} - k_{\text{dRaf1}}[\text{Raf1PPtase}]\frac{[\text{Raf1Active}]}{[\text{Raf1Active}]+K_{\text{mdRaf1}}} - k_{\text{dRaf1ByAkt}}[\text{AktActive}]\frac{[\text{Raf1Active}]}{[\text{Raf1Active}]+K_{\text{mRaf1ByAkt}}}$$

$$\frac{d[\text{BRafInactive}]}{dt} = -k_{\text{Rap1ToBRaf}}[\text{Rap1Active}]\frac{[\text{BRafInactive}]}{[\text{BRafInactive}]+K_{\text{mRap1ToBRaf}}} + k_{\text{dBRaf}}[\text{Raf1PPtase}]\frac{[\text{BRafActive}]}{[\text{BRafActive}]+K_{\text{mdBRaf}}}$$

$$\frac{d[\text{BRafActive}]}{dt} = +k_{\text{Rap1ToBRaf}}[\text{Rap1Active}]\frac{[\text{BRafInactive}]}{[\text{BRafInactive}]+K_{\text{mRap1ToBRaf}}} - k_{\text{dBRaf}}[\text{Raf1PPtase}]\frac{[\text{BRafActive}]}{[\text{BRafActive}]+K_{\text{mdBRaf}}}$$

$$\frac{d[\text{MekInactive}]}{dt} = -k_{\text{pRaf1}}[\text{Raf1Active}]\frac{[\text{MekInactive}]}{[\text{MekInactive}]+K_{\text{mpRaf1}}} - k_{\text{pBRaf}}[\text{BRafActive}]\frac{[\text{MekInactive}]}{[\text{MekInactive}]+K_{\text{mpBRaf}}} + k_{\text{dMek}}[\text{PP2AActive}]\frac{[\text{MekActive}]}{[\text{MekActive}]+K_{\text{mdMek}}}$$

$$\frac{d[\text{MekActive}]}{dt} = +k_{\text{pRaf1}}[\text{Raf1Active}]\frac{[\text{MekInactive}]}{[\text{MekInactive}]+K_{\text{mpRaf1}}} + k_{\text{pBRaf}}[\text{BRafActive}]\frac{[\text{MekInactive}]}{[\text{MekInactive}]+K_{\text{mpBRaf}}} - k_{\text{dMek}}[\text{PP2AActive}]\frac{[\text{MekActive}]}{[\text{MekActive}]+K_{\text{mdMek}}}$$

$$\frac{d[\text{ErkInactive}]}{dt} = -k_{\text{pMek2}}[\text{MekActive}]\frac{[\text{ErkInactive}]}{[\text{ErkInactive}]+K_{\text{mpMek2}}} + k_{\text{dErk}}[\text{PP2AActive}]\frac{[\text{ErkActive}]}{[\text{ErkActive}]+K_{\text{mdErk}}}$$

$$\frac{d[\text{ErkActive}]}{dt} = +k_{\text{pMek2}}[\text{MekActive}]\frac{[\text{ErkInactive}]}{[\text{ErkInactive}]+K_{\text{mpMek2}}} - k_{\text{dErk}}[\text{PP2AActive}]\frac{[\text{ErkActive}]}{[\text{ErkActive}]+K_{\text{mdErk}}}$$

$$\frac{d[\text{PI3KInactive}]}{dt} = -k_{\text{PI3K}}[\text{boundEGFR}]\frac{[\text{PI3KInactive}]}{[\text{PI3KInactive}]+K_{\text{mPI3K}}} - k_{\text{PI3KRas}}[\text{RasActive}]\frac{[\text{PI3KInactive}]}{[\text{PI3KInactive}]+K_{\text{mPI3KRas}}}$$

$$\frac{d[\text{PI3KActive}]}{dt} = +k_{\text{PI3K}}[\text{boundEGFR}]\frac{[\text{PI3KInactive}]}{[\text{PI3KInactive}]+K_{\text{mPI3K}}} + k_{\text{PI3KRas}}[\text{RasActive}]\frac{[\text{PI3KInactive}]}{[\text{PI3KInactive}]+K_{\text{mPI3KRas}}}$$

$$\frac{d[\text{AktInactive}]}{dt} = -k_{\text{Akt}}[\text{PI3KActive}]\frac{[\text{AktInactive}]}{[\text{AktInactive}]+K_{\text{mAkt}}}$$

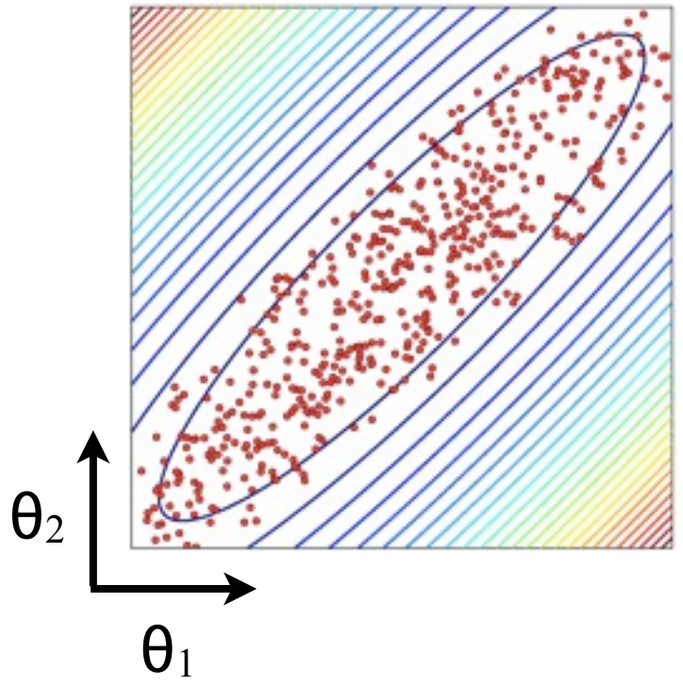
$$\frac{d[\text{AktActive}]}{dt} = +k_{\text{Akt}}[\text{PI3KActive}]\frac{[\text{AktInactive}]}{[\text{AktInactive}]+K_{\text{mAkt}}}$$

$$\frac{d[\text{C3GInactive}]}{dt} = -k_{\text{C3GNGF}}[\text{boundNGFR}]\frac{[\text{C3GInactive}]}{[\text{C3GInactive}]+K_{\text{mC3GNGF}}}$$

$$\frac{d[\text{C3GActive}]}{dt} = +k_{\text{C3GNGF}}[\text{boundNGFR}]\frac{[\text{C3GInactive}]}{[\text{C3GInactive}]+K_{\text{mC3GNGF}}}$$

$$\frac{d[\text{Rap1Inactive}]}{dt} = -k_{\text{C3G}}[\text{C3GActive}]\frac{[\text{Rap1Inactive}]}{[\text{Rap1Inactive}]+K_{\text{mC3G}}} + k_{\text{RapGap}}[\text{RapGapActive}]\frac{[\text{Rap1Active}]}{[\text{Rap1Active}]+K_{\text{mRapGap}}}$$

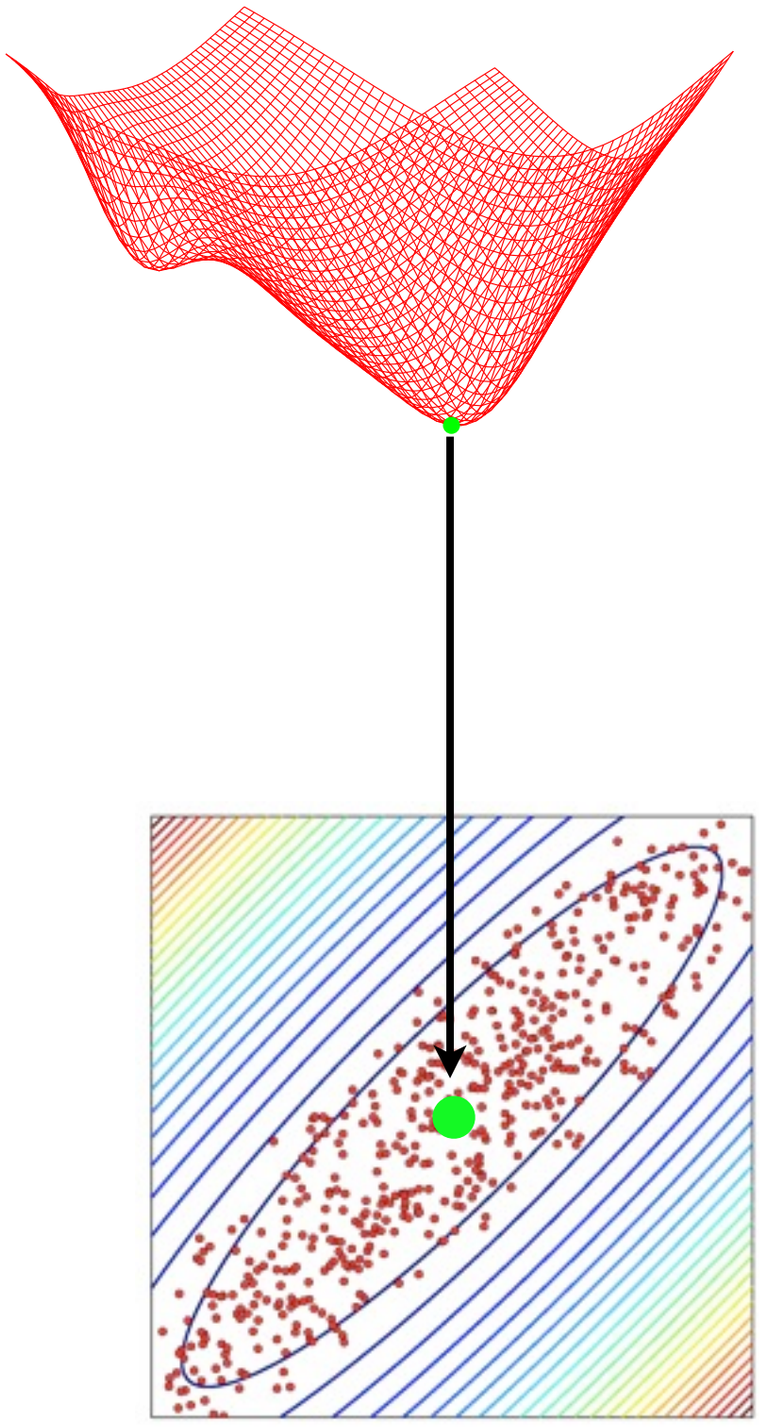
$$\frac{d[\text{Rap1Active}]}{dt} = +k_{\text{C3G}}[\text{C3GActive}]\frac{[\text{Rap1Inactive}]}{[\text{Rap1Inactive}]+K_{\text{mC3G}}} - k_{\text{RapGap}}[\text{RapGapActive}]\frac{[\text{Rap1Active}]}{[\text{Rap1Active}]+K_{\text{mRapGap}}}$$



$$\frac{d[\text{SosActive}]}{dt} = +k_{\text{EGF}}[\text{boundEGFR}]\frac{[\text{SosInactive}]}{[\text{SosInactive}]+K_{\text{mEGF}}} + k_{\text{NGF}}[\text{boundNGFR}]\frac{[\text{SosInactive}]}{[\text{SosInactive}]+K_{\text{mNGF}}} - k_{\text{dSos}}[\text{P90RskActive}]\frac{[\text{SosActive}]}{[\text{SosActive}]+K_{\text{mdSos}}}$$

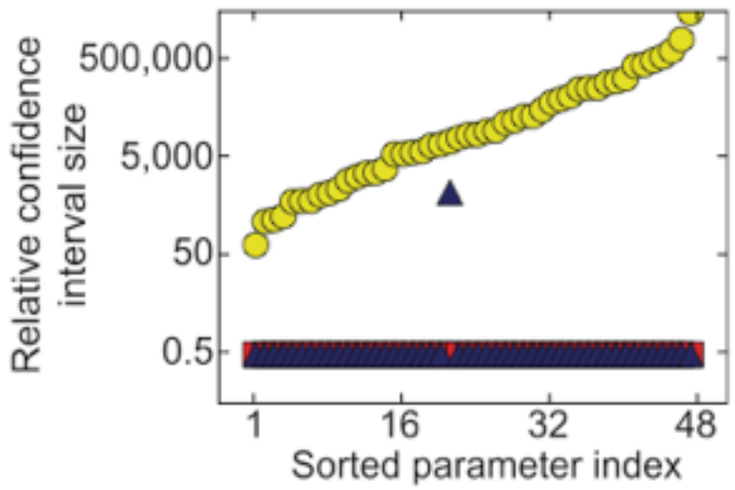
How does the behavior of a model change with changes in model parameters?
(Implications for model validation, experimental design, evolution, control, etc.)

Fits, ensembles, and uncertainties

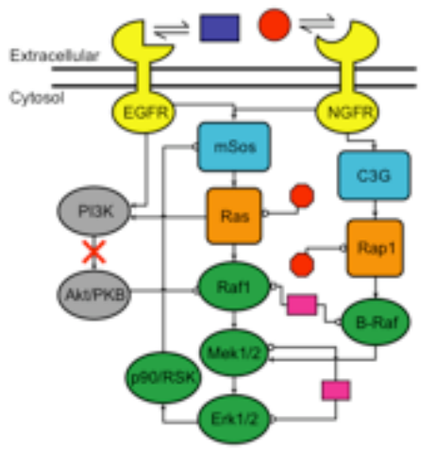


Chemical species $y(\vec{\theta},t)$
Parameters $\vec{\theta}$
Data $y_i \pm \sigma_i$
Cost $C(\vec{\theta})$

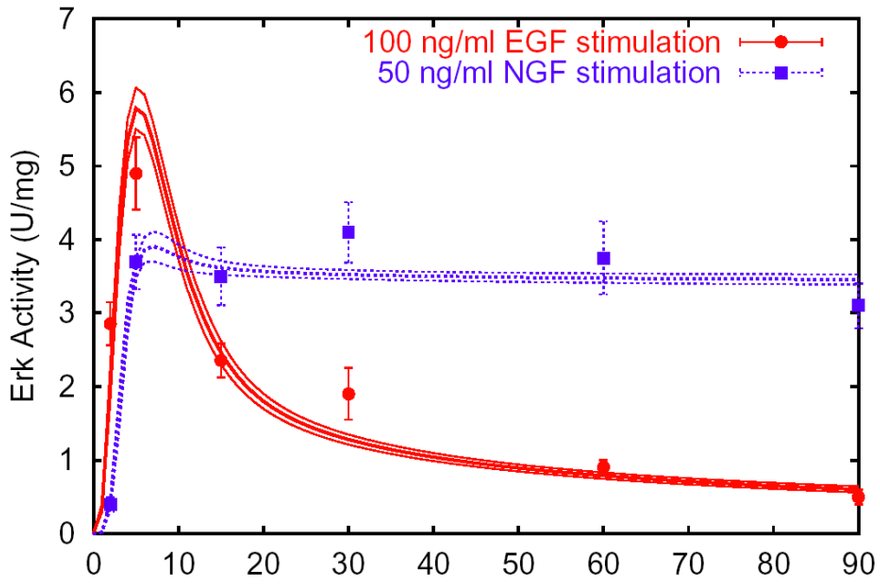
$$C(\vec{\theta}) = \frac{1}{2}\sum_{i=1}^{N_D} \frac{(y(\vec{\theta}) - y_i)^2}{\sigma_i^2}$$



tight predictions despite underlying parametric uncertainty

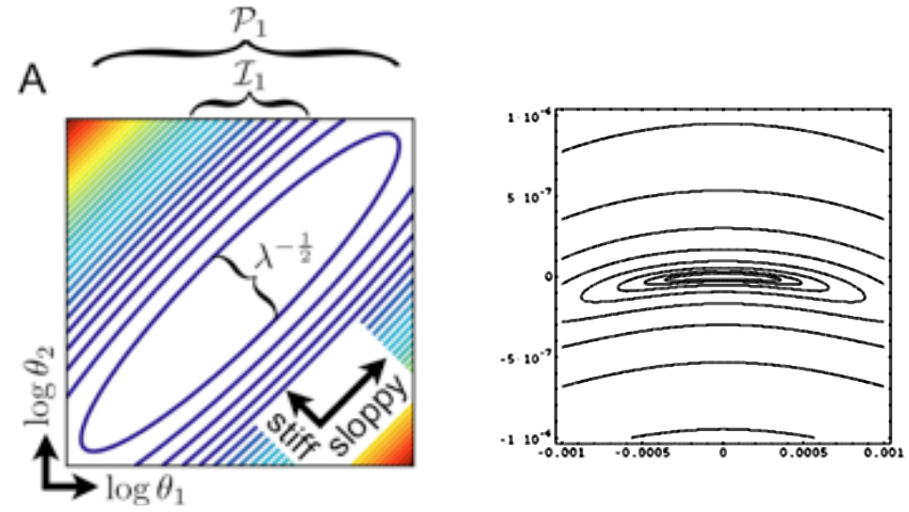


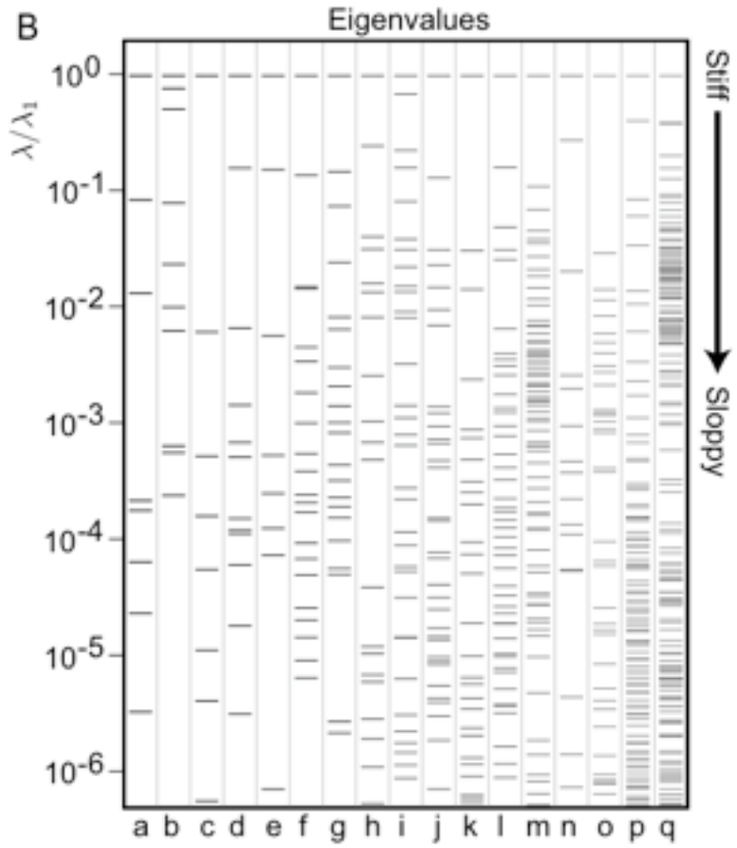
Prediction uncertainties (from linearized analysis or sampling of parameter ensembles)



Sloppy parameter sensitivities

- Behavior of model is typically orders of magnitude more sensitive to moves along some directions in parameter space than others
  - "stiff" vs. "sloppy"
- Characterized by the eigenvalues of the Hessian matrix about the best fit
  - eigenvalues span many orders of magnitude (ill-conditioned)
  - roughly constant density in log λ: *a few stiff modes, and many sloppy ones*
  - similar for χ^2 analysis of "complete data" generated by model with reference parameters
  - ubiquitous across systems biology models
- Implications for experimental design
  - don't focus on parameter identification

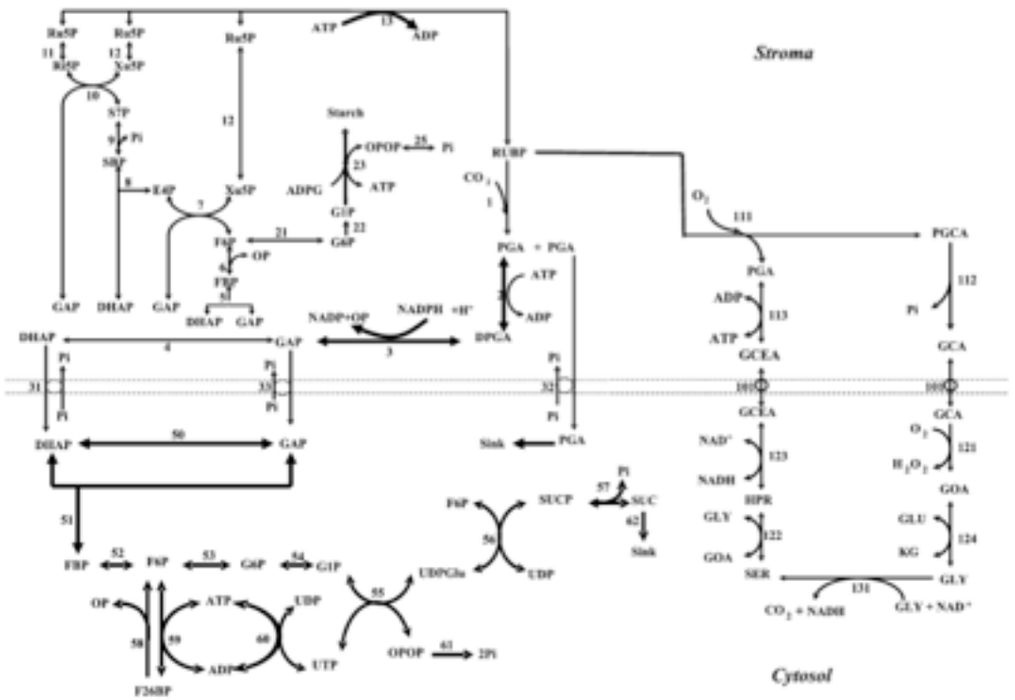




(a) Tyson 1991
(b) Zwolak 2005
(c) Goldbeter 1991
(d) Vilar 2002
(e) Edelstein 1996
(f) Kholodenko 2000
(g) Lee 2003
(h) Leloup 1999
(i) Brown 2004
(j) von Dassow 2000
(k) Ueda 2001
(l) Locke 2005
(m) Zak 2003
(n) Curto 1998
(o) Chassagnole 2002
(p) Chen 2004
(q) Sasagawa 2005

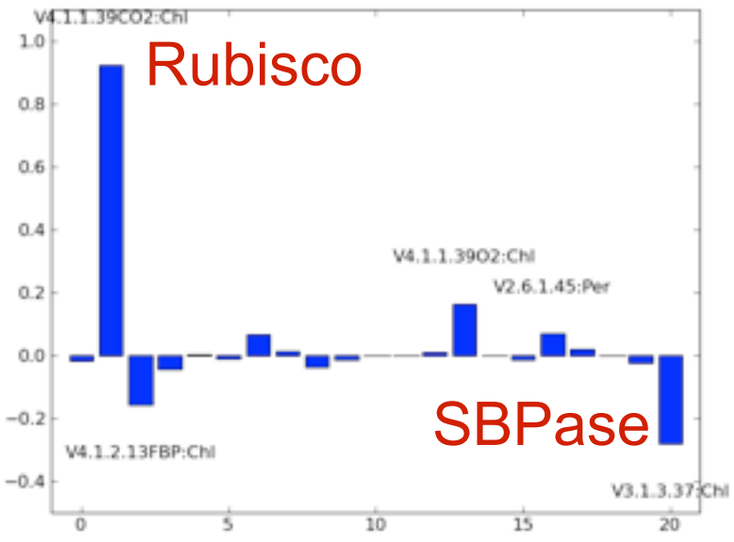
R.N. Gutenkunst *et al.*, PLoS Comp. Bio. (2007)

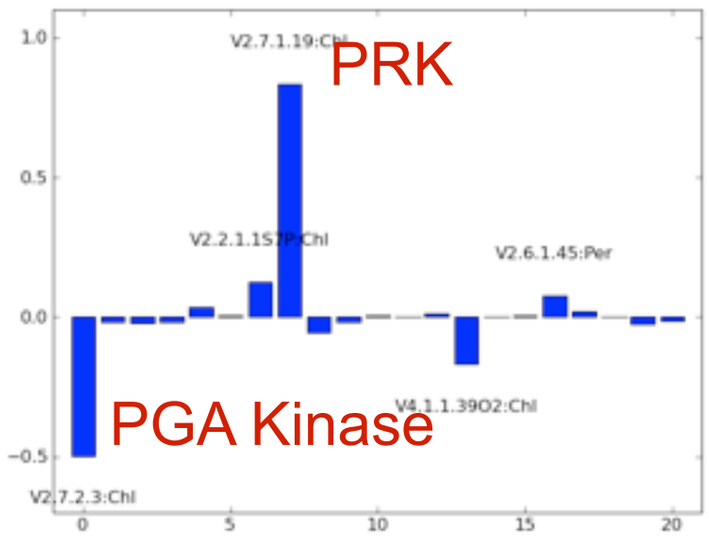
Evolvability of photosynthetic metabolism



- C3 photosynthesis model (Zhu 2007): evolutionary optimization of CO2 assimilation by reallocating enzyme concentrations
- examining “stiffness” and “sloppiness” of collective parameter sensitivities to probe structure of evolutionary path
- need for visualization tools to see dynamics and sensitivity analysis in context of pathway diagram
- need for tools/formats for model & simulation exchange and validation (a la BioModels.net)

three “stiffest” parameter directions from preliminary sensitivity analysis (“Eigen Plot”)

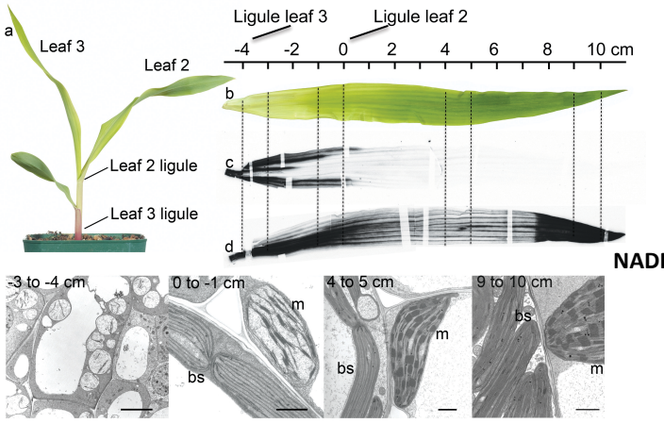




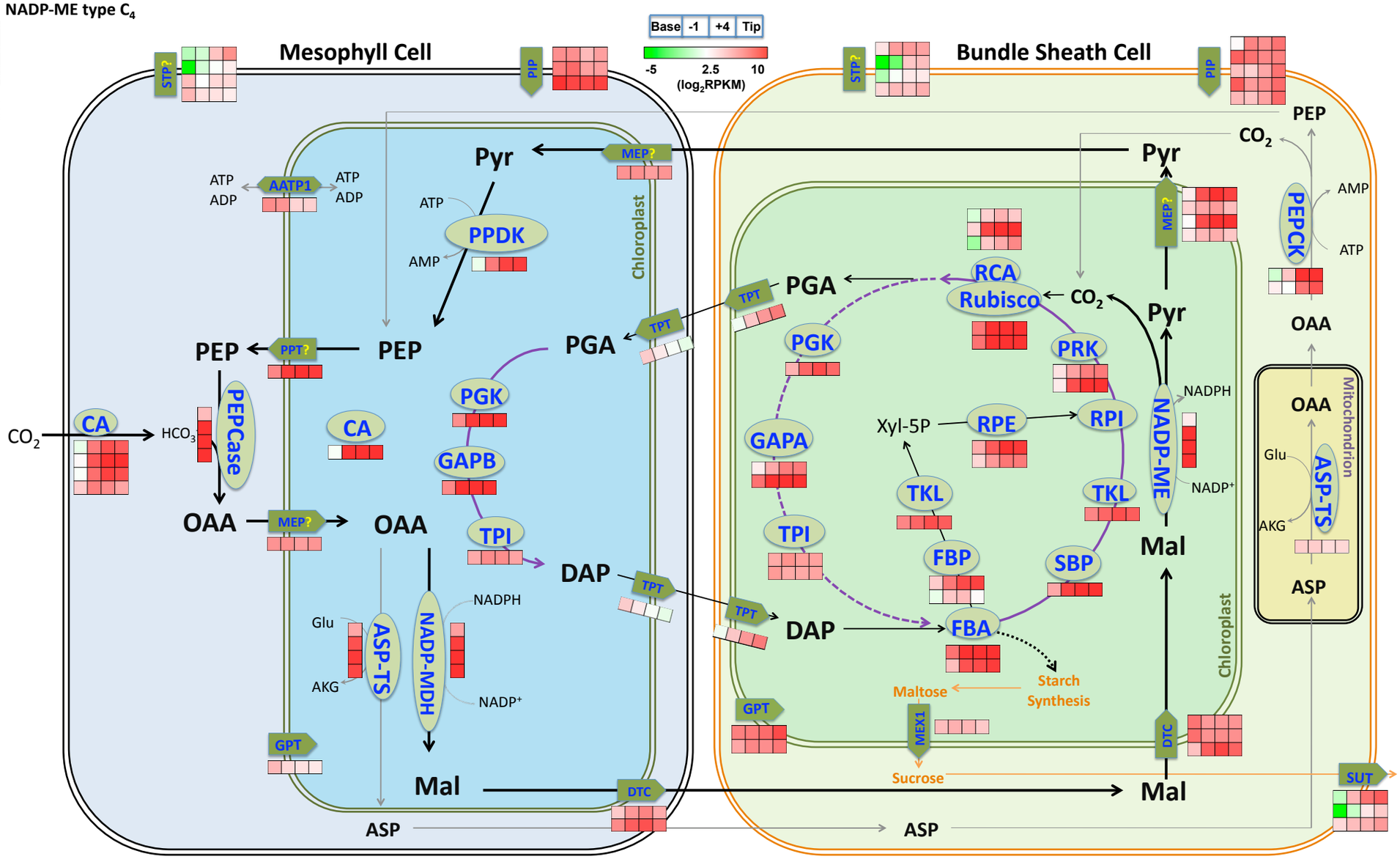
SloppyCell

- open-source software system for kinetic modeling, parameter estimation, uncertainty quantification and sloppy modeling
  - sloppycell.sourceforge.net; built largely on Python & SciPy (for numerics)
- SBML read & write: allows for analysis of a broad range of published models
- symbolic engine for synthesizing model ODEs and parametric sensitivity equations from network topology and kinetic expressions
- multiple algorithms for least-squares parameter fits to data (gradient-based, gradient-free, information geometry)
- sensitivity analysis (linearized, and via Monte Carlo sampling of parameter space)
- SloppyCell in its entirety, or in pieces carved off as separate libraries, are available for further development in iPlant

Pathways & visualization



NADP-ME type C_4



PPDK: Pyruvate Orthophospho-Dikinase
PRP: PPDK – regulatory protein
PEPCase: Phospho*enol*Pyruvate Carboxylase
CA: Carbonic Anhydrase
NADP-MDH: NADP – malate dehydrogenase
TPT: Phosphate/triose-phosphate translocator

NADP-ME: NADP – Malic enzyme
PPT: Phosphate/PhosphoenolPyruvate translocator
DTC: dicarboxylate/tricarboxylate carrier
MEP: Envelope protein
PEPCK: PhosphoenolPyruvate carboxykinase
Rubisco: Ribulose-1,5-bisphosphate carboxylase/oxygenase

RCA: Rubisco activase
PRK: Phosphoribulokinase
RPI: ribose-5-phosphate isomerase
TKL: Transketolase
SBP: Sedoheptulose-1,7-bisphosphatase
FBA: Fructose-bisphosphate aldolase

PGK: Phosphoglycerate kinase
TPI: Triosephosphate isomerase
GAPA: Glyceraldehyde-3-phosphate dehydrogenase A
GAPB: Glyceraldehyde-3-phosphate dehydrogenase B
ASP-TS: Aspartate Aminotransferase
DAP: dihydroxyacetone phosphate

Standards & formats for modeling

from BioModels.net

Standard specification of quantitative models	*Model description*	*Simulation description*	*Simulation results description*
Minimal requirements	MIRIAM	MIASE	
Data format	SBML, SBGN	**SED-ML**	**SBRML**
Ontologies	SBO	KiSAO	TEDDY

- MIRIAM: Minimum Information Required in the Annotation of Models
- MIASE: Minimum Information About a Simulation Experiment
- SBML: Systems Biology Markup Language
- SBGN: Systems Biology Graphical Notation
- SED-ML: Simulation Experiment Description Markup Language
- SBRML: Systems Biology Results Markup Language
- SBO: Systems Biology Ontology
- KiSAO: Kinetic Simulation Algorithm Ontology
- TEDDY: Terminology for the Description of Dynamics

Proposal: Devote iPlant resources to determine if:
(a) these standards are useful in supporting model validation, comparison and coupling,
(b) there are existing tools, or a need for further tool development, to support effective use of these standards and formats,
(c) these standards are applicable beyond chemical kinetics, e.g., to ecophysiology (or epidemiology)